उत्तर प्रदेश राजर्षि टण्डन मुक्त
विश्वविद्यालय, इलाहाबाद

UGSTAT-12 / PGSTAT-22 / MASTAT-22

Printed Chandrakala Universal Pvt Ltd, 42/7 J.L.N. Road Allahabad

National and International Statistical System

Indian Statistical System

Historical perspective

In India, collection of Statistical data which were of use to sovereign was practiced even in ancient times. More recently, the Mughals had a system of collection and compilation of crop statistics to help them in land revenue collection. With the progress of British rule these systems died out. The British, in course of time, created
their own data-generating systems to serve their specific ends. These systems of data collection did not develop into an integrated or well-coordinated statistical system during colonial rule and their coverage was largely limited to a few specific fields like trade and commerce, production of selected industries, population, some basic crop statistics and livestock.

Independence in 1947 ushered in an era of economic planning and emphasized the necessity of a strong database covering a variety of social and economic topics. The system of data collection left behind by the colonial rulers as found to be far from adequate to meet this immediate demand. It did not even provide the basic data required for estimation of national income which was so essential for assessing the performance and progress of the economy. Data reflecting the conditions of the vast majority of the population languishing in a state of chronic hunger was hard to come by. There was no provision for generating employment statistics for large section of the population engaged in agriculture and other unorganized sectors of the economy. The immediate task, therefore was to set up a statistical system capable of filling large data gaps in the data essential for formulating effective economic plans.

At the instance of Prime Minister Jawaharlal Nehru, a modest beginning was made in 1949. The National Income Committee was appointed by the Government of India to work out a reliable method of estimating national income. On its recommendation, the Directorate of National Sample Survey was set up in 1950 to collect essential statistics relating to socio-economic characteristics and agricultural production. In the following year, the Central Statistical Organisation was formed. In short, the foundations of a modern statistical system was laid in those early years of independence. Since then, the sustained efforts of academics and official statisticians have seen the collection of statistics in India gradually mature into one of the most comprehensive statistical system of the developing world.

Present Indian Statistical System

The Indian Statistical System functions within the overall administrative framework of the country. The Indian federal structure has influenced the organization of the statistical system as well. The division of administrative functions between the Government of India and the State Governments is on the basis of the subject classifications under the Union, State and Concurrent Lists as detailed in the Constitution of India. At the Centre, the responsibilities are further divided amongst the various ministries and departments, according to the Allocation of Business Rules, 1961 that are amended from time to time. The collection of statistics on any subject generally vests in the authority (Central Ministry or Department or State Government Department) that is responsible for that subject according to its status in the Union, State or Concurrent Lists. By and large, the flow of statistical information emanates from the States to the Centre except in cases where the State level operations are an

integral part of Centrally sponsored schemes or data are collected through national sample surveys.

Statistical System at the Centre

The collection of statistics for different subject-specific areas, like agriculture, labour, commerce, industry etc. vests with the corresponding administrative ministries. More often than not, the statistical information is collected as a by-product of administration or for monitoring the progress of specific programmes. Some of the ministries, like Agriculture, Water Resources, Health etc. have full-fledged statistical divisions, while most others have only a nucleus cell. Large-scale statistical operations like the Population Census, Annual Survey of Industries, Economic Census, etc. are generally centralized, and these cater to the needs of other ministries and departments, as well as State Governments. In important ministries, officers of the Indian Statistical Service (ISS) and subordinate statistical staff perform the statistical functions. The Central Statistical Office (CSO) in the Ministry of Statistics and Programme Implementation (MoS&PI) is the nodal agency for a planned development of the statistical system in the country and for bringing about coordination in statistical activities among statistical agencies in the Government of India and State Directorates of Economics and Statistics.

Statistical System in the States

The Statistical System in the States is similar to that at the Centre. It is generally decentralized laterally over the Departments of the State Governments, with major Departments, such as, Agriculture or Health, having large statistical divisions for the work of departmental statistics. At the apex level is the Directorate (formerly Bureau) of Economics and Statistics (DES), which is formally responsible for the coordination of statistical activities in the State. The DESs have large organizations at the headquarters, with statistical offices in the districts and, in some cases, in the regions of the State. The statistical activity of the DESs is more or less uniform. They publish statistical abstracts and handbooks of the States, annual economic reviews or surveys, district statistical abstracts, and State budget analysis; work out the estimates of the State Domestic Product and Retail Price Index Numbers and engage in such other statistical activities as is relevant to the State. Most of them participate at least on a matching sample basis in the national Sample Survey Programme, and some of them carry out an Annual Survey of Industries (ASI) for factories not covered by the ASI of the National Sample Survey Office (NSSO). Generally, the States do not have a common statistical cadre. Some of the states have made common Statistical cadre.

System Flow

Indian Statistical System is a system built upwards from village to block to district to State Government departments, and from there to the corresponding ministries at the Centre. This is its true representation: a collection of State-level systems forming a National system. In the current context, making the National system the starting point, from the perspective of an inverted view the system can be described as laterally decentralized among the Ministries of the Government of India (GOI), and in every one of them, vertically decentralized, between the Centre and the States. The bond between the State departments and the Central Ministries in the field of statistics has always been strong.

The question of lateral coordination, among Ministries at the Centre and among the Departments in the States, was not much of an issue when statistics were compartmentalized. It is a later-day need when, for treating all Government statistics as one subject, Central

statistical offices were created at the Centre (the CSO) and in the States (the State Directorates of Economics and Statistics). When a single office was charged with the responsibility of bringing together all statistics relating to the country, and of examining them from the perspectives of quality, timeliness, accuracy and other desirable merits which statistics should possess, a new situation was created which, although it did not make the Ministry strictly answerable or accountable to the CSO for its statistics, required it at least to share information on the methodology and procedures of collecting those statistics with an office outside of itself. This created a new requirement, that of effective lateral coordination between the CSO on the one hand and the ministries on the other, for the system to function satisfactorily. The same situation was created in most States where also the system is fully laterally decentralized.

The main features of the Indian Statistical System can be thus summarized as :

(a) The Administrative Statistics System is its major component;

(b) It is laterally and vertically decentralized;

(c) In it, not only data collection but also compilation, processing and preparation of results are carried out by the States for most of the sectors; and

(d) It is the State-wise results, which flow to the Centre, and statistics at the all-India level are obtained as the aggregates of State-level statistics.

International Statistical System

The International Statistical System is, by and large, decentralized by subjects. The UN Statistical Division (UNSD) is entrusted with the task of minimum coordination of related activities. The efforts of the UNSD and other international agencies are directed towards providing conditions under which improved and comparable national statistics can be produced to serve national as well as international purposes. The international statistical cooperation is carried out with the help of many international agencies. The article 55 of the Charter of the UN provides for creation of Economic and Social Council. The Council is responsible for the economic and social activities of the UN. It makes recommendations to the General Assembly, member countries and specialized agencies and works through its subsidiary organs namely, 5 Regional Commissions, 9 Functional commissions and 14 UN Specialized Agencies, as given in Annexure.

. The statistical activities of the UN are carried out under the guidance of the Statistical Commission which advises both the Council and the Secretary General. The International Statistical Organisations (UN Statistical Division, UN Statistical Commission, Regional Commissions and Specialised Agencies) help to improve and expand the Statistical operations of many developing countries by providing (i) advisory services, (ii) training facilities, (iii) technical guidance in preparation and application of guidelines, manuals, etc., and (iv) funding of certain national programmes.

Regional Commissions

The Regional Commissions have been established to enable the nations of the major regions of the world to cooperate on common problems and also to produce economic and social information. In creating the Regional Commissions, the Council has given each one considerable latitude in devising its own rules or procedures and in specifying the significant spheres of operation and concentration of its resources. Each Regional Commission has established a statistical division. The work programmes of these statistical divisions are determined not only by the Regional Commissions but also by the UN Statistical Commission. India is a member country of the Economic and Social Commission for Asia and the Pacific (ESCAP) region.

The functions of the Regional Commissions are as under :

(i) the introduction and application of world-wide standards within the region in order to achieve better regional and international comparability of data;

(ii) the development of national statistical services, with the provisions of various forms of assistance to the Governments concerned, if necessary;

(iii) the discussion of various methodological, organizational, data processing and other technical questions of special interest in the region;

(iv) the dissemination of information on experiences gained within and outside the region with respect to new statistical concepts and techniques for the benefit of all the countries participating in the Commission;

(v) undertaking statistical activities for the Commission or one of its organs to be used in economic surveys and other studies and providing other statistical services of a general nature.

UN Statistical Commission

The Statistical Commission plays a significant role in the coordination of international statistical activities not only within the UN System, but also in relation to other international bodies. So far as possible and practicable, statistical relationship between the UN and specialized agencies is reciprocal in respect to the obligations assumed and benefits derived. The specialized agencies have the right and obligation to collect and use statistics, which uniquely pertain to the functions of each.

The Statistical Commission has 24 members who are appointed in their individual capacity by the UN Secretary General with the approval of the national Governments. The representation from the States of various regions is as under : African States- 5, Asian States- 4, Eastern European States-4, Latin American & Carribbean States- 4, and Western Europe and other States-7. The term of office of each member is 4 years.

The Commission assists the Council in :

- promoting the development of national statistics and the improvement of their comparability;
- coordinating the international statistical activities not only within the UN, but in relation to other international bodies;
- the development of the central statistical services of the Secretariat;
- advising the organs of the UN on general questions relating to the collection, interpretation and dissemination of statistical information; and
- promoting the improvement of statistics and statistical methods generally.

Specialised Agencies

The Specialised Agencies are separate, autonomous organizations related to the UN by special arrangements. They work with the UN and with each other through the coordinating machinery of the Council. However, these specialized agencies report their activities to the Council.

The technical assistance activities of the Specialized Agencies include all major fields of statistics. These activities are divided among the organizations of the UN System in accordance with their overall areas of responsibility. For example, technical assistance in the field of Labour Statistics is dealt with by the ILO; Agricultural, Forestry and Fisheries Statistics by FAO; Health Statistics by WHO; Educational, Cultural and Scientific Statistics by UNESCO; Civil Aviation Statistics by ICAO; Monetary, Balance of Payments and

Banking Statistics by International Monetary Fund (IMF); and demographic, social statistics, foreign trade, industrial and construction statistics, price statistics, national accounts, transport and energy statistics and data processing etc. by the UNSD.

Technical Assistance in Statistics

The ways in which technical assistance in statistics is rendered are similar. There are however, differences in the emphasis given to the various forms of assistance between the members of the UN family. The main forms of technical assistance are as follows :

(a) Advisory services (experts; regional and inter-regional advisors),
(b) Training (international statistical training centers, fellowships for training, programmes at the headquarters or regional offices of international organisations; ad-hoc training course, workshops and seminars),
(c) Provision of equipment,
(d) Technical guidance (Preparation, publication and application of international guidelines, technical manuals and other documents, working groups and other technical meetings),
(e) Programme formulation, support and evaluation (assessment of the national, regional and inter-regional requirements, aid in the recruitment and technical supervision or technical assistance, experts).

UN Secretariat

The UN Secretariat has assumed such statistical obligations and functions including those of coordination as are inherent in its central position, among the specialized agencies under the UN Charter.

UN Statistical Division

The main responsibility of the UNSD located in the UN Secretariat is to serve the General Assembly and its different political bodies. The UNSD has been entrusted with the following functions : (i) to promote the development of national statistics and the improvement of international comparability, and (ii) to serve as an international center for the collection, compilation, analysis, evaluation and publication of statistics, so that data in each field can be made readily available for both international and national use. The UNSD is also recognized as the central agency for publication of statistics.

The UNSD has special responsibility for the overall coordination of the above activities, and in that regard, special importance is attached to the following :

(a) the avoidance of unnecessary reporting burden on national authorities, and
(b) the coordination of statistical standards and methods.

Coordination at the international level is done by the Statistical Commission which makes the initial recommendations to the UNSD for the development of new/improvement of existing statistics in a particular field. The UNSD prepares the draft proposals either from its own resources (manpower) or by assigning the work to experts or expert bodies. In this process, it consults the Regional Commissions, Specialised Agencies and the international professional bodies like International Statistical Institute (ISI). The draft proposals are sent to the member countries for their comments and suggestions. The comments received from the countries are coordinated and new draft proposals are framed for consideration by the Statistical Commission at its next session. In case the Commission rejects the proposal, it gives specific directions for further improvement in the standard. The standard is again

modified and the revised version is prepared for further consideration, by the Commission at its next session. Finally when the Commission is satisfied with the proposals, which have been submitted, these are put up in the form of recommendations to the Economic and Social Council. The Council makes recommendations by passing resolutions to the National Governments for implementation of its proposals.

List of UN Regional Commissions, Functional Commissions and Specialized Agencies under Economic and Social Council

I. Regional Commissions:

1. The Economic Commission for Europe (ECE)
2. The Economic & Social Commission for the Asia and the Pacific (ESCAP)
3. The Economic Commission for Africa (ECA)
4. The Economic Commission for Latin America and the Caribbean (ECLAC)
5. The Economic and Social Commission for Western Asia (ESCWA)

II. Functional Commissions:

1. Statistical Commission
2. Commission on Population and Development
3. Commission for Social Development
4. Commission on Human Rights
5. Commission on Narcotic Drugs
6. Commission on the Status of Women
7. Commission on Crime Prevention and Criminal Justice
8. Commission on Science and Technology for Development
9. Commission on Sustainable Development

III. Specialized Agencies

1. International Labor Organization (ILO)
2. Food & Agriculture Organization of the United Nations (FAO)
3. United Nations Educational, Scientific & Cultural Organization (UNESCO)
4. International Civil Aviation Organization (ICAO)
5. World Health Organization (WHO)
6. Universal Postal Union (UPU)
7. International Telecommunication Union (ITU)
8. World Meteorological Organization (WMO)
9. International Maritime Organization (IMO)
10. World Intellectual Property Organization (WIPO)
11. World Bank Group (IBRD, IDA, IFC, MIGA, ICSID)
12. International Monetary Fund (IMF)
13. International Fund for Agriculture Development (IFAD)
14. United Nations Industrial Development Organization (UNIDO)

IV. United Nations Forum on Forests

Sessional and Standing Committees. Expert, ad-hoc and related bodies

Functions and Activities of Ministry of Statistics and Programme Implementation and Central Statistic Office

The Ministry of Statistics and Programme Implementation (MoS&PI) has two wings, one relating to Statistics and the other Programme Implementation. The Statistics Wing of MoS&PI consists of the Central Statistics Office (CSO), National Sample Survey Office (NSSO), and Computer Centre. The programme Implementation wing has four Divisions, namely (i) Twenty Point Programme (ii) Infrastructure Monitoring (iii) Project Monitoring and (iv) Member of Parliament Local Area Development Scheme. The Central Statistics Office has 5 divisions; (i) National Accounts Division (NAD), (ii) Training Division, (iii) Social Statistics Division (SSD), (iv) Economics and Statistics Division (ESD) and (v) Coordination and Administration Division (CAD).

The Statistics Wing of MOS&PI is the apex body in the official statistical system of the country and is mandated the following responsibilities:

(i) Acting as the nodal agency for planned development of the statistical system in the country.

(ii) Coordinating statistical work with a view to identifying gaps in data availability or duplication of statistical work in respect of the Departments of the Government of India and State Statistical Bureaus (SSBs) and suggesting necessary remedial measures.

(iii) Maintaining liaison with international statistical organisations, such as, the United Nations Statistical Division (UNSD), the Economic and Social Commission for Asia and the Pacific (ESCAP), the Statistical Institute for Asia and the Pacific (SIAP), the International Monetary Fund (IMF), the Asian Development Bank (ADB), the Food and Agriculture Organisation (FAO), the International Labour Organisation (ILO), etc. in different statistical matters which cover interchange of statistical intelligence, participation in international conferences, seminars, workshops for the development of capabilities for improving statistical system and serving as the clearing house for statistical matters.

(iv) Laying down and maintenance of norms and standards in the field of statistics, involving concepts and definitions, methodology of data collection, processing of data and dissemination of results.

(v) Advising the Ministries/Departments of the Government of India on statistical methodology and on statistical analysis of data.

(vi) Preparing national accounts as well as publication of annual estimates of national product, government and private final consumption expenditure, capital formation, saving, estimates of capital stock and consumption of fixed capital, as also state level gross capital formation of supra-regional sectors and preparing comparable estimates of State Domestic Product (SDP) at current prices.

(vii) Compiling and releasing the Index of Industrial Production (IIP) every month in the form of 'quick estimates'; conducting the Annual Survey of Industries (ASI); and providing statistical information to assess and evaluate the changes in the growth, composition and structure of the organised manufacturing sector.

(viii) Organising and conducting periodic all-India Economic Censuses and Follow-up Enterprise surveys.

(ix) Conducting large scale all-India sample surveys for creating the database needed for studying the impact of specific problems for the benefit of different population groups in diverse socio-economic areas, such as employment, consumer expenditure, housing conditions and environment, literacy levels, health, nutrition, family welfare, etc.

(x) Examining the survey reports from the technical angle and evaluation of appropriate sampling design including survey feasibility studies in respect of surveys conducted by the National Sample Survey Organisation and other Central Ministries and Departments.

(xi) Providing an in-house facility to process the data collected through various socio-economic surveys and follow-up enterprise surveys of Economic Censuses conducted by the National Sample Survey Organisation and the Central Statistical Organisation as well as providing the computing services to other agencies/ individuals.

(xii) Disseminating statistical information on various aspects through a number of regular, or ad hoc publications distributed to Government, semi-Government, or private data users/ agencies; and dissemination of data, on request, to the United Nations agencies like the UNSD, the ESCAP, the ILO and other international agencies.

(xiii)Giving grants-in-aid to registered Non-Governmental Organizations and research institutions of repute for undertaking special studies or surveys, printing of statistical reports, and financing seminars, workshops and conferences relating to different subject areas of official statistics.

(xiv) Functioning as the Cadre Controlling Authority and dealing with the aspects of managing the Indian Statistical Service and Subordinate Statistical Service including matters pertaining to skill upgradation, career progression and manpower planning.

(xv) Acting as the nodal Ministry for the Indian Statistical Institute and ensuring its functioning in accordance with the provisions of the Indian Statistical Institute Act, 1959 (57 of 1959).

The Programme Implementation Wing of the Ministry has been entrusted with the following responsibilities:-

(i) Monitoring of the Twenty Point Programme (TPP);

(ii) Monitoring the performance of the country's eleven key infrastructure sectors, viz., Power, Coal, Steel, Railways, Telecommunications, Ports, Fertilizers, Cement, Petroleum & Natural Gas, Roads and Civil Aviation;

(iii) Monitoring of all Central Sector Projects costing Rs.20 crore and above; and

(iv) Monitoring the implementation of the Member of Parliament Local Area Development Scheme (MPLADS).

Besides, the MOSPI has been entrusted with the responsibility of Coordination and Policy issues relating to the National Common Minimum Programme (NCMP) excluding sectoral policies allocated to other Ministries/Departments.

CENTRAL STATISTICS OFFICE

The Central Statistics Office (CSO) was set up in May 1951 with the purpose of coordination of statistical activities of the different Ministries of the Government of India and the State governments and for promotion of statistical standards. Since then the CSO has been coordinating the Statistical activities in the country, including laying down and maintenance of statistical norms and standards and providing liaison with Central, State and International statistical agencies. The CSO also shoulders the responsibility of preparation of Accounts of National Income, conduct of Annual Survey of Industries, Economic Censuses and their follow up surveys, compilation of Index of Industrial Production, Consumer Price Indices for Urban

Non-Manual Employees, Human Development Statistics and Gender Statistics, imparting training in Official Statistics, Five Year Plan Work relating to Development of Statistics in the States and Union Territories, Dissemination of Statistical Information, work relating to Trade, Energy, Construction and Environment Statistics, revision of National Industrial Classification, etc.

The Director General is the head of the organization who is assisted by 5 Additional Director Generals, 13 Deputy Director Generals, Directors and other officers/supporting staff. The CSO is located at New Delhi. However, the work pertaining to Annual Survey of Industries (ASI) is carried out at the Industrial Statistics Wing of the CSO (CSO-IS Wing), located at Kolkata.

Coordination of Statistical Activities

In fulfilling its coordinating role, the CSO participates in the inter-departmental meetings of the Working Groups, Technical Advisory Committees, Standing Committees constituted by the Central Ministries/ Departments in subject areas of their concern with a view to ensuring adoption of statistical standards, minimizing duplication of efforts and for promoting quality and timeliness of data. The CSO also participates in the meetings of the high-level statistical coordination committees of the State Governments, which review activities related to statistical data generation and dissemination.

To strengthen the coordination of statistical activities among the Ministry of Statistics and Programme Implementation, other Central Ministries and State Statistical Organizations, the CSO organizes the Conference of Central and State Statistical Organizations (COCSSO).

Earlier it was organized once in every two years. Keeping in view the usefulness of the Conference, it has now been decided to have it every year. A Standing Committee under the chairpersonship of Director General, CSO takes follow up action on the recommendations/ suggestions made by the COCSSO.

The CSO provides secretarial assistance to the National Advisory Board on Statistics (NABS), the apex body set up by the Government of India to provide an overall perspective for statistical development and to identify priority areas. The division-wise activities are detailed in the following paragraphs-

National Accounts

The National Accounts System intends to present a unified statistical picture of the economy with details of inter-relations among sectors, among different type of activities and among regions, as are required for understanding the complete behaviour of the economy. The CSO prepares national accounts, as well as publishes annual estimates of national product, government/ private final consumption expenditure, capital formation and saving along with consolidated accounts of the nation and detailed accounts of the public sector transactions. The other activities undertaken are:

(i) preparation of quarterly estimates of gross domestic product (GDP) at current and constant prices;

(ii) estimation of capital stock and consumption of fixed capital;

(iii) estimation of state level gross capital formation of supra-regional sectors of railways, communication, banking and insurance, and central government administration; and

(iv) preparation of comparable estimates of state domestic product (SDP) at current prices for use of the Finance Commission and the Planning Commission.

The annual publication 'National Accounts Statistics' gives estimates of the national /domestic product, private final consumption expenditure, saving, capital formation, capital stock, public sector transactions and consolidated accounts of the nation.

As per the recommendations of National Statistical Commission, the CSO is to provide technical leadership, guidance and co-ordination in the compilation of National/ Regional Accounts to the States. In pursuance of the above recommendation, the CSO conducts regional training-cum-workshops on regional accounts for states.

Industrial Statistics

The work relating to Industrial Statistics falls under three categories viz. (i) Annual Survey of Industries, (ii) Index of Industrial Production, and (iii) Standardization of Industrial Classifications.

Annual Survey of Industries

The Annual Survey of Industries (ASI) is the principal source of industrial statistics in India. It provides statistical information to assess and evaluate, objectively and realistically, the change in the growth, composition and structure of the organised manufacturing sector comprising activities relating to manufacturing processes, repair services, generation and transmission of electricity, gas and water supply and cold storage. The survey is conducted annually under the statutory provisions of the Collection of Statistics Act, 1953. The ASI covers the entire country except the States of Arunachal Pradesh, Mizoram, Sikkim and Union Territory of Lakshadweep. It covers all factories registered under Sections 2m(i) and 2m(ii) of the Factories Act, 1948. The survey also covers bidi and cigar manufacturing establishments registered under the Bidi and Cigar Workers (Conditions of Employment) Act, 1966. All the electricity undertakings engaged in the generation, transmission and distribution of electricity, registered with the Central Electricity Authority were covered under the ASI, irrespective of their employment size up to the ASI 1997-98. This has, however, been discontinued from ASI 1998-99 and onwards. Certain services and activities like cold storage, water supply, repair of motor vehicles/ consumer durables like watches, etc., are covered under the survey. Defence establishments, oil storage and distribution depots, restaurants, hotels, cafe, computer services and the technical training institutes are excluded from the purview of the survey.

The data collected through the ASI relate to capital invested, employment, emoluments, labour cost, consumption of fuel and lubricants, raw material, total input/ output, gross value added, net value added and other characteristics of factories/ industrial establishments. The Field Operations Division (FOD) of the NSSO carries out the fieldwork. The CSO processes the data and publishes the results annually through detailed reports and also makes them available electronically.

Index of Industrial Production

Index of Industrial Production (IIP) is an up-to-date indicator of industrial growth in the Indian economy. The all India IIP is a composite indicator that measures the short-term changes in the volume of production of a basket of industrial products during a given period with respect to that in a chosen base period. It is compiled and published monthly by the Central Statistics Office (CSO) with the time lag of six weeks from the reference month. In other words, IIP is an abstract number, the magnitude of which represents the status of production in the industrial sector for a given period of time as compared to a reference period of time. It is a statistical device which enables us to arrive at a single representative figure to measure the general level of industrial activity in the economy. Strictly speaking, the IIP is a short-term indicator of industrial

growth till the actual results from Annual Survey of Industries (ASI) and National Accounts Statistics become available. This indicator is of paramount importance to the Government for policy planning purposes and is also being extensively used by various organizations including Industrial Associations, Research Institutes and Academicians.

The base of current series of Index of Industrial Production (IIP) is 2004-05. This series covers 682 items (clubbed in 399 item groups: Mining-1, Manufacturing – 397, Electricity – 1) comprising 61 from Mining and Quarrying, 620 from Manufacturing and 1 from Electricity Sector having the weightage of 141.57%, 755.57% and 103.16% respectively in the all-India IIP. The IIP is released every month in the form of Quick Estimates with a time lag of less than 6 weeks as per the SDDS norms of the IMF. The estimates are also simultaneously being released as per use-based classification viz., Basic Goods, Capital Goods, Intermediate Goods, Consumer goods, Consumer Durables and Consumer Non-durables. These estimates get revised subsequently on receipt of updated production data from the 15 different source agencies in various Ministries/ Departments/ Attached/ Subordinate Offices. However, the major source of data is the DIPP, Ministry of Industry and Commerce, which provides data for 209 item groups contributing 52.0% weight in the overall IIP.

Standardisation of Industrial Classification

The National Industrial Classification(NIC), released by the CSO, is a classification of economic activities undertaken by the economic units. It does not draw the distinction according to the kind of ownership, type of legal organisation, type of technology and scale/ mode of operation. The NIC was last updated in 2004. It follows the principles of the latest UN System of Industrial Classification for all Economic Activities, 2002 (ISIC Revision 3.1). The NIC-98 had adopted ISIC Rev. 3 up to 4-digit level in toto and extended them to the 5th digit to describe more detailed activities at the national level. In NIC-2004, the NIC-98 has been updated with a limited objective of accommodating national requirements while maintaining international comparability with the updated ISIC Rev.3.1. The exercise does not effect any major changes in the structure of the existing classification i.e. NIC-98 except removing ambiguity/duplications/omissions etc. by suitably modifying the explanatory notes. The major structural changes required in the classification will be considered in the next revision of NIC in pursuance of new version of ISIC Rev.4 to be released in 2007.

Economic Census

The CSO has been making concerted efforts to develop a sound and reliable database on unorganized non-agricultural economic sectors. To achieve this, a scheme 'Economic Census and Follow-up Surveys' was launched in 1976 to fill the data gaps in the unorganized segments of the economy. The Economic Census (EC) provides basic details about all enterprises in the country. The follow up surveys provide detailed information relating to various parameters associated with the enterprise. This information is used by different category of users and also for preparation of National Accounts.

After the first EC in 1977, the second and third ECs were conducted alongwith the house-listing operations of the 1981 and 1991 Population Censuses. The fourth EC was conducted independently by the CSO in collaboration with the State Directorates of Economics and Statistics, covering establishments and own account enterprises of the entire economy except those engaged in crop production and Plantation. The fifth EC was held during Marc 2005 – March 2006.

The follow-up surveys are the only source of information on the unorganized sector of non-agricultural production for the preparation of National accounts. The CSO has been conducting sample suryes on different sub-sectors of the unorganized sectors.

Social Statistics

The CSO plays the role of a coordinator for the development of social statistics, which, *inter-alia*, cover population, human development, employment, health, education, social justice and women empowerment etc. Close liaison is maintained with different agencies like the Planning Commission, Office of the Registrar General of India, Ministries of Human Resource Development, Labour, Urban Development and Poverty Alleviation, Social Justice and Empowerment, Health and Family Welfare besides autonomous research institutions like International Institute of Population Sciences (IIPS), Indian Council of Social Science Research (ICSSR), National Institute of Public Cooperation and Child Development (NIPCCD), National Council of Educational Research and Training (NCERT), Institute of Applied Manpower Research (IAMR), National Council of Applied Economics Research (NCAER), etc., for effectively discharging responsibilities. The CSO also shares its expertise with various organizations in the field of social statistics by participating in the meetings of the Expert Groups/ Working Groups, Technical Advisory Committees and other Committees constituted by different ministries/ departments/ organizations and providing with expert advisory support on the proposals, questionnaire, tabulation programme, draft reports etc., received from them. The CSO publishes ' Selected Socio-Economic Statistics, India' which gives time series data on different social indicators.

The Social Statistics Division is responsible for coordinated development of social statistics which inter-alia covers population, human development, employment, health, education, social justice, women empowerment, disability, environment, gender statistics and statistical monitoring of Millennium Development Goals & SAARC Social Charter Goals in addition to implementation of India Statistical Strengthening Project, and the scheme of awards and fellowships for outstanding and meritorious research work in statistics.

Millennium Development Goals (MDGs): The Central Statistics Office is vested with the responsibility of Statistical Monitoring of Millennium Development Goals (MDGs).The CSO in coordination with the participating Ministries has been tracking the MDGs, which have been internalized with the policy and programme initiatives of the country. The Second India - specific country report on MDGs. viz. 'Millennium Development Goals, India Country Report, 2007', brought out during the year captures inter-alia data/information pertaining to India's challenges, achievements, policies and programmes from both the national and sub-national perspectives. India's approach to the MDG statistics has received attention in the region.

From 1975, the International Year of Women, there has been a growing interest in gender statistics. In India, this is one of the most important emerging areas that have received attention of the policy makers and all other concerned professionals in recent years. The need for reliable statistical information on gender issues, which are required for understanding the role of women in the society, has been felt with increasing urgency. For some areas of interest, like population, mortality, literacy and workforce, gender specific statistics have been in the public domain for quite some time. For the others, the statistics from various sources needs to be collected. The CSO took steps to implement an ESCAP project on Improvement of Statistics on Gender Issues during 1994 to 1996. The outcome of these efforts was a publication entitled "Women and Men in India", brought out for the first time by the CSO in 1995. The publication is being brought out regularly.

SAARC Gender Data Base: The CSO is the lead agency to facilitate development of SAARC Gender Data Base. A prototype of the data base has been developed along with list of statistical indicators.

SAARC Social Charter: The Ministry of Statistics and Programme Implementation has been designated as the nodal Ministry in respect of the SAARC Social Charter signed by the SAARC Heads of States at the 12th Summit. The Secretary, Ministry of Statistics and Programme Implementation is the Chairman of the National Coordination Committee (NCC) constituted to facilitate the implementation of the SAARC Social Charter and monitor the performance in achieving the goals.

Environment Statistics

The CSO has assumed responsibility for development of statistical standards in environment statistics and strengthening of database in this area. A small cell has been created in the CSO for compilation of data on environment statistics under the six internationally recognized heads: Flora, Fauna, Atmosphere, Water, Land/Soil and Human Settlements. The CSO brings out a publication 'Compendium of Environment Statistics' which gives a comprehensive picture of environmental degradation, its causes and the areas of future concern.

Price Statistics

The CSO compiles and publishes, on monthly basis, Consumer Price Index for Urban Non-Manual Employees [CPI(UNME)] since 1961. Presently the CPI with base 1984-85 is being compiled. Data on urban retail prices for items and services in the consumption basket, used for compilation of CPI(UNME), are collected by the Field Operations Division (FOD) of the National Sample Survey Organisation (NSSO) from 59 centres for about 250 items, including 15 selected service items. The number of goods and services for which data are collected varies from centre to centre. The smallest number is 146 in the case of Imphal; and the largest 345 for Delhi. Data on house rent and off-take in respect of selected commodities are also collected. This index is widely used by various agencies for varied purposes including as an input to the wage and salary adjustment process; indexation of pension and superannuation payments, Government taxes, charges and contracts; and as a general measure of inflation. However, CPI(UNME) is would be discontinued after the proposed series viz. CPI(Urban) is brought out.

International Comparison Programme

The International Comparison Programme (ICP) developed by the United Nations facilitates comparison of macro- economic aggregates like GDP, GNP, etc. between different countries in a more meaningful way (by using Purchasing Power Parity [PPP] numbers) than the comparison based on the Official Exchange Rates. So far seven rounds of ICP (starting from 1970 to 2005) have been conducted and India has participated in all the rounds except the sixth round. India participated in the 7th Round of ICP which had 2005 as the reference year. The role of the participating countries was limited to collection of prices as per the specifications prescribed and furnishing the same along with basic heading weights from GDP expenditure estimates. Actual compilation of PPP numbers was done by the coordinating agencies in respective regions. Collection of prices of items identified under ICP (7th Round) was undertaken in three phases (Food, Clothing & Footwear; Other household consumer items and educational & medical services) by the National Sample Survey Organization (Field Operations Division) during April, 2005 to June, 2006. In case of items pertaining to construction and plant & machinery, price data were collected with the cooperation of concerned Government Ministries/ Departments.

The final results of the 7th Round of ICP undertaken during 2005 giving PPPs of the currencies of participating economies of the Asia - Pacific region have been released by the Asian Development Bank, the Coordinating agency for this region. The preliminary global

results of the 7th Round of ICP undertaken during 2005 have also been released by the World Bank.

India Statistical Strengthening Project (ISSP)

The Ministry of Statistics and Programme Implementation (MOSPI) is the participating
entity in respect of a World Bank assisted project 'India Statistical Strengthening Project' (ISSP) for strengthening of Indian Statistical System. The ISSP has adopted a 2 - Tier approach for implementation. In Tier - I, those activities were taken up for study which were necessary to provide adequate baseline details for the objective and data based formulation of Tier – II activities, besides providing some operationally /functionally useful results. MOSPI now has available fairly detailed and comprehensive State/UT-Specific Study Reports on 'Identifying the Specific Requirements for Strengthening of the State Statistical Bureaus', for 35 states/UTs. At present, the process of synthesis of the State/UT reports is under way.

A Concept Paper for Tier - II of the project approved by the National Statistical Commission and the Planning Commission has been posed to the World Bank. The ISSP, besides supporting the project implementation within the MOS&PI, envisages to extend support
to the State/UT Governments in primarily the following identified areas.

(i) Improving the coordination and management of statistical activities in the States;
(ii) Human Resource Development (HRD);
(iii)Development of Statistical Infrastructure;
(iv)Investments in Physical Infrastructure, including IT and Improving Statistical Operations, especially those that contribute to the cause of improvement of Quality and dissemination of statistical data.

Human Resource Development

Training is the most important tool of Human Resource Development (HRD) in any organization. The Training Division of the Central Statistical Organisation (CSO) is mandated with the task of looking after the training needs of not only ISS officers but also of all statistical personnel in the other Central Government Departments, States/UTs and Public Sector Undertakings in the country in addition to participants from abroad by providing need based training in the field of Official Statistics and related disciplines. The Training Division is headed by an Additional Director General supported by a Deputy Director General and a team comprising of Director/Joint Director/Deputy Director level officers of Indian Statistical Service (ISS) besides others belonging to subordinate statistical service. With a view to meet the increasing challenges faced by the statistical systems and to equip the personnel with latest techniques of statistical data management, the Ministry has decided to transform this Division into a National Academy of Statistical Administration (NASA) which is likely to be operational
during 2008-09 at Greater Noida, Gautam Budh Nagar, Uttar Pradesh with state-of-the-art facilities.

A high level body "The Training Programmes Approval Committee" (TPAC) comprising senior officers of the Indian Statistical Service (ISS) acts as an advisory body in designing the need-based training programmes and also periodically reviews, inter alia, the syllabi and training methodology for various courses conducted by the Training Division.

Data Dissemination

The CSO collects, compiles and updates various socio-economic data series and disseminates data to various national and international agencies, including UNSD, ESCAP, ILO and ADB. The following publications are being regularly brought out:

(i) Monthly Abstract of Statistics – Monthly (Bilingual)
(ii) Statistical Abstract, India – Annual (English)
(iii) Statistical Pocket Book, India – Annual (English)
(iv) India in Figures – Annual (Bilingual)

Apart from the above regular publications, the CSO also brings out the following publications:

(a) National Accounts Statistics -Sources and Methods
(b) Directory of Sample Surveys in India
(c) Guide to Official Statistics
(d) Directory of Statistics, India
(e) Statistical System in India
(f) Selected Socio-Economic Statistics

Origin and functions of NSSO

The need for developing a sound database for launching planning era in respect of various fields was keenly felt by Late Pandit Jawaharlal Nehru, the first Prime Minister of free India as early as in 1948. It was, at his instance and on the recommendations of the National Income Committee chaired by Late Professor P.C.Mahalanobis (other members, Dr. C.D.Deshmukh and Prof. V.K.R.V. Rao), a large scale sample survey agency known as NSS (National Sample Survey) came into existence in 1950.While instituting the NSS, the Government of India turned to the Indian Statistical Institute for providing technical direction to the surveys. The work relating to finalization of sampling design, schedules of enquiries, writing of instructions, training of field staff, processing of data and writing of reports was all entrusted to the Indian Statistical Institute.

NSS Directorate

A field branch of the survey named as Directorate of NSS was set up under the Department of Economics in the Ministry of Finance, Government of India and assigned the job of conducting the fieldwork. Shri S.P. Sinha was appointed as the first chief Director of the NSS Directorate.

The first round of NSS, which was carried out during October 1950-March 1951 was devoted tocollection of data on consumer expenditure and employment and unemployment conditions in the Country besides other related data. The Pune based Gokhale Institute of Poltics and Economics was also involved in the design of the first survey. The first round survey had four sets of schedules: (A): Village schedule for collecting data like land utilization, prices of selected items, wage rates etc. (B): household schedule for collection of demographic particulars, land holding, employment etc. (C): household enterprises schedules for collecting information on household enterprises and allied activities: while the last set (D) household consumption schedule for collecting data on household consumption on various items.

Prof. P.C. Mahalanobis played a crucial role in the establishment of NSS. Besides looking after the entire technical work of the survey ISI also undertook the field work directly in West Bengal and Bombay. The direct involvement of ISI in NSS continued up to 1970. However there were inordinate delays in the processing and publishing of reports. Government of India thereafter reviewed the functioning of the NSS. The review Committee under the then Cabinet Secretary Shri B. Sivaraman (other members- Prof. V.M.Dandekar and Prof. R.R.Bahadur) proposed the complete merger of all the aspects of the survey work under a single organization. This organization was named as the National Sample Survey Organization (NSSO) and was placed under the Department of Statistics and now with Ministry of Statistics &P.I .of the Government of India.

Reorganisation

In March 1970, the NSS was reorganized and all aspects of its work were brought under a single Govt. organisation, namely the National Sample Survey Organisation (NSSO) under the overall technical guidance of Governing Council to impart objectivity and autonomy in the matter of collection, processing and publication of the NSS data.

Consequent to the setting up of the National Statistical commission(NSC), the governing council has been dissolved with effect from 30th August, 2006 as all the functions of the Governing Council has been assumed by the NSC. The NSSO is mandated to conduct nation-wide surveys on various socio-economic aspects, Annual Survey of Industries follow up surveys of Economic Census and supervision of area enumeration and crop estimation surveys of state agencies under the Improvement of Crop Statistics Scheme, Urban Frame Survey, organization of Methodological studies and pilot surveys on important subjects and collection of price data from rural and urban sectors

Steering Committee for National sample Surveys

The NSC in its 3rd meeting recommended for the constitution of a Steering Committee for National Sample Surveys and its composition. Accordingly, the government has constituted the Steering Committee for National Sample Surveys under the Chairmanship of Prof. S.D. Tendulkar. The Committee comprises of 7 Non-official members and 8 officials. Director General &CEO, NSSO is the Convenor of the Steering Committee.

The terms of reference of the Steering Committee are:

(a) Make recommendations to the NSC in respect of the following:
   (i) Short term and long term programme for National Sample Surveys including topics of surveys, periodicity and subjects to be recovered.
   (ii) Methodological improvements in the conduct of National Sample Surveys.

(b) Finalise the sample design, concepts and definitions, questionnaires, tabulation plans, etc. for specific round of Sample Survey.

(c) Prepare studies for improving survey methodology, data collection, processing and dissemination and any other issue as may be referred to it by the Commission.

(d) Approve the survey reports for release.

(e) May constitute Expert Groups for any specific technical issue that requires expertise beyond the core competency of the Steering Committee.

Survey Cycle

NSSO has been conducting multi-subject integrated household sample surveys since 1950. The subject coverage of the socio-economic surveys conforms to a well-defined cycle of the surveys extending over a period of 10 years. Surveys on consumer expenditure and employment & unemployment, social consumption (health, education etc.), manufacturing enterprises and service sector enterprises in the unorganised sectors are covered once in five years, while subjects like land holdings & livestock holdings and debt & investment are covered once in 10 years. Thus, out of a cycle of 10 years, pre-assigned subjects are allocated for nine years while one year is kept as an open round to cover special topics of current interest to meet the demand of the data users. The data on consumer expenditure and employment and unemployment are also collected in every round from a small sample of households along with the main survey of enquiry. These socio-economic surveys are conducted in the form of rounds, the period of a round being normally a year. Subject coverage in the last 10 rounds is given in Annex I

NSSO also undertakes the fieldwork of Annual Survey of Industries under statutory provisions of the Collection of Statistics Act, 1953 (Central Rules, 1959), covering all the factories registered under Sections 2m(i) and 2m(ii) of the Factories Act, 1948, which refer to the establishments using power and employing 10 or more workers and those not using power and employing 20 or more workers on any day of the preceding 12 months and bidi and cigar units (principal employer)registered under the Bidi and Cigar Workers (Conditions of Employment) Act, 1966.

NSSO also provides technical guidance to states in the field of agricultural statistics for conducting crop estimation surveys and keeps a continuous watch on the quality of crop statistics through the Improvement of Crop Statistics scheme.

NSSO regularly collects retail prices from shops/outlets in a sample of 603 villages for compilation of Consumer Price Index numbers for rural labour. Till March 2008, Price Data was collected from selected markets of 59 urban centres for compilation of Consumer Price Index Numbers for Urban Non –Manual Employees. With effect from 1st April 2008, price data is being collected for the base year of Consumer Price Index (Urban) from 310 towns.

NSSO conducts Urban Frame Survey (UFS) for providing sampling frame of first stage units in the urban sector.

The NSSO is headed by the DG&CEO. The organisation has four divisions, namely, Survey Design and Research Division (SDRD), Field Operations Division (FOD), Data Processing Division (DPD) and Coordination & Publication Division (CPD). A Deputy Director General heads each Division except FOD. An Additional Director General heads FOD.

Survey Design and Research Division

SDRD is located at Kolkata. It is responsible for

- Planning of the Survey
- Formulation of sampling design

- Formulation of Concepts and definitions
- Drawing up of survey schedules
- Writing of instructions
- Preparation of validation and tabulation programmes
- Finalisation of survey results and release of reports
- Providing technical guidance in sampling techniques to various official agencies.

Field Operations Division

FOD has its headquarters at New Delhi and Agricultural Wing at Faridabad. It has 6 Zonal offices located at Bangalore, Kolkata, Guwahati, Jaipur, Lucknow and Nagpur, 49 Regional offices and 116 sub-regional offices located in different parts of the country.

Functions of FOD are:

- Collection of data through (i) Annual Survey of Industries (ii) Socio-economic surveys, (iii) Price collection surveys, (iv) Follow up surveys of Economic census
- Updation of Urban Frame Survey blocks
- Sample check on area enumeration and crop cutting experiments and providing technical guidance to the states for improvement of crop statistics
- Providing in-service training to field staff
- Adhoc surveys (important recent surveys- Non profit Institutions and serving households, UNICEF survey on wellbeing of women and children)

Data Processing Division

DPD with its headquarters at Kolkata is responsible for processing of survey schedules canvassed in the various socio-economic surveys. The Data Processing activities have been decentralised through its six centres located at Ahmedabad, Bangalore, Kolkata, Delhi, Giridih and Nagpur. The various functions of DPD are

- Sample selection and preparation of sample list.
- Preparation/checking of multiplier details
- Pre-data entry scrutiny of filled-in-schedules
- Development of software
- Data entry and verification
- Scrutiny of error list which include content checking - missing records, duplicate records and inconsistent records
- Coverage check to ensure completeness of the data file
- Preparation of trial and final tables
- Software assistance to State authority for processing of state sample data

Coordination & Publication Division

CPD has its headquarters at New Delhi and is responsible for

- Coordinating the activities of all the four divisions of NSSO
- Dissemination of survey results and analysis through the biannual technical journal "Sarvekshana".

- Providing assistance to the Governing Council of NSSO
- Supplying survey data of various rounds to individuals, researchers, research Institutions and other private and govt. bodies
- Liaison with other Departments/Ministries on various matters concerning NSSO.

Dissemination of survey results and data

Results of NSSO surveys are brought out in the form of NSS reports. So far, 526 reports have been brought out. NSS reports are available for sale. Summary Results are also published in Sarvekshana, a biannual technical journal of NSSO.

Validated unit level data relating to various surveys of NSSO are available on magnetic media (CD- ROM) for sale at nominal price.

MOU with research institutes and universities

The approved research institutes and universities in India and abroad can obtain NSS data free of cost for research/studies concerning national development and planning by signing a Memorandum of Understanding (MOU) with the National Sample Survey Organisation.

What is Statistics? At the outset, it may be noted that the word 'Statistics' is used in two senses-plural and singular. In the plural sense it is used to mean numerical data arising in any sphere of human experience. It is in this sense that the term is used when we talk about the production and sale of textile or television sets or average rainfall in a region and so on. Used as singular, 'Statistics' is a name for the body of scientific methods which are meant for collection, analysis and interpretation of numerical data. According to Spiegal, Statistics is concerned with scientific method for collecting, organizing, summarizing, presenting and analyzing data as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis.

Statistical data: Statistics being a body of methods meant for the study of numerical data, it is obvious that the first step in any statistical enquiry must be collection of relevant data. Data may be of two broad types: primary and secondary. Any data that have been collected earlier for some other purpose are secondary data in the hands of an individual who is using them. In contrast, the data that are collected first hand by someone specifically for the purpose of facilitating the study are known as primary data. These are collected by the enquirer, either on his own or through some agency set up for the purpose, directly from the field of enquiry. Thus, the primary data collected by some person may become the secondary data for another. For example, the demographic statistics collected every ten years are the primary data with the Registrar General of India, but the same used by anyone else would be secondary data with that individual.

The main advantage of secondary data is that it is far more economical as the cost of collecting original data is saved and it saves considerable amount of time in comparison with primary data which incurs much more time and cost.

The main advantage of primary data is that it can always be used with greater confidence, because the enquirer himself can decide upon the coverage of data and the definitions to be used and, as such, have a measure of control on the reliability of the data. In making use of secondary data one has to be particularly careful about the nature of the data-their coverage, the definitions on which they are based and their degree of reliability.

Variables and Attributes: Regardless whether the data are primary or secondary, they can be classified in a different way viz. variables and attributes. Quantitative data that can be expressed in numerical terms are called variables whereas qualitative data that can not be expressed in numbers are called attributes. For example, height, weight, income, expenditure of an individual or a group of individuals can be expressed in numbers and hence they are variables. On the other hand, sex of a newborn baby, languages people speak in a region etc. can not be expressed in numeric terms and hence they are attributes.

Quantitative data or variable again may be of two principal types. If the variable takes only some isolated values, like the number of letters in a word, number of members in a family and so forth, it is called a discrete variable. On the other hand, if the variable takes any value within its range of variation, it is called a continuous variable. The height of a man, the diameter of a bobbin, the rainfall or humidity in a region are examples of continuous variables. Although in case of continuous variables actual measurements present a discreteness, e.g. when heights are given correct to the nearest Centimetre (cm), this discreteness is completely artificial, being due to the limitations of the measuring instruments.

Presentation of statistical data: After data collection the data the next important job is the presentation of the collected data in a comprehensive and readily understandable

manner. There are broadly three methods of presenting statistical data – textual, tabular and diagrammatic presentation of data.

One of the common methods of presenting numerical data is to use paragraphs of text. This method is almost invariably employed in official reports and in reports sent to management, either alone or in combination with the tabular and diagrammatic methods. The textual mode of data presentation has an appeal to people with a literary bent of mind, who have distaste for the drabness of a table, and to those who are not satisfied with the broad trend of the data that a graph or some other diagram may reveal. An obvious disadvantage of this method is that it is not very useful when a large mass of data is to be presented or when a large number of comparisons are to be made or attention of the reader is to be drawn to a large number of points. A long textual presentation may turn out to be boring and monotonous for the reader and can even dissociate the reader. Thus care should be taken to make such presentation objective and unambiguous. It should be brief and precise and should follow a logical sequence.

Another important method of presentation of data is in the form of a table. The basic purpose of tabulation is to condense the mass of data and to make it understandable. Presentation of data by means of tables are generally preferred because a table can show the data in a compact form and a complete table with its title, heading and footnote can bring all the essential features of the data into a clearer perspective. It also enables comparisons to be easily made and attention to be readily drawn to important features of the data. While presenting data in tabular format it should be kept in mind that the table should be well balanced in length and breadth and also the arrangement of items should follow a logical sequence. The table should be complete in all respects including the units of measurement, the time period to which the data relate, abbreviations used, footnote and source note, if necessary.

The third important method of presenting statistical data is by means of diagrams-by graphs, charts or pictures. On account of their visual impact, the data presented through diagrams have a ready appeal and are better grasped and remembered than the tabulated data. Being readily intelligible, diagrams are almost essential whenever it is required to convey any statistical information to the general public. It should be remembered, however, that information on a limited number of topics only can be presented in a single diagram so as to maintain its neatness and keep it simple. When too many details are to be presented, these devices fail to do so without loss of clarity. Moreover, a diagram can give only a rough idea about the magnitude of the variation, whereas in a table the exact values may be quoted. Thus diagrammatic presentation can not be a complete substitute of a tabular or textual presentation; rather, they can serve the purpose better when accompanied by suitable tables or substantiated by a proper text. The more important types of diagram which are used in statistical work are Line diagrams and Bar diagrams.

Line diagrams are generally used to present time-series data i.e. data varying over time. Time (in suitable units, say years) is generally taken along the X-axis and the variable under study (e.g. yield of wheat in a particular state or industrial growth in the economy) along the Y-axis. The units and scales are suitably chosen and the line diagram is drawn by joining the points plotted on a graph.

Bar diagrams can be used for series varying either over time or over space. In this method, bars (rectangles) of equal width are taken for the different items of the series, the length of a bar representing the value of the variable concerned. Conventionally, horizontal bars are used to represent spatial series while vertical bars are used for series varying over time.

Apart from line and bar diagrams, pictorial diagrams are also used to present numerical data where a suitable symbol is first chosen to represent a certain number of units of variable. Each value of the given series is then represented by taking a number of symbols (including a fraction of a symbol) of the same size.

In cases where we are interested more in the percentages for different categories (i.e. in their relative importance) rather than in their absolute values, Pie diagrams are used. In pie diagram a circle is used, the area enclosed by it being taken as 100. It is then divided into a number of sectors by drawing angles at the centre, the area of each sector representing the corresponding percentage. Thus 360 degree at centre represents 100% and hence for any particular category the angle is 3.6 times the corresponding percentage. (See Annex for examples of Line, Bar and Pie diagrams.)

Frequency Distribution: When observations are available on a single characteristic for a large number individuals, often it becomes necessary to condense the data as far as possible without losing any information of interest. Let us consider the following example.

Example1.1 Suppose the marks obtained in Statistics by 50 students in an examination are available and they are given in Table 1.1

Table 1.1: Marks obtained in Statistics by 50 students.

32	55	67	74	45	67	54	85	39	78
25	62	55	83	67	48	36	74	32	45
74	39	67	55	54	32	83	78	28	55
45	85	25	32	48	62	74	54	36	28
58	58	74	83	67	48	36	45	78	83

This representation of data does not furnish any useful information and is rather confusing. A better way may be to express the figures in an ascending or descending order of magnitude commonly termed as an array. But this does not reduce the bulk of the data; a much better representation is given in the following table.

Table 1.2: Frequency distribution of individual marks.

Marks	Number of students	Marks	Number of students	Marks	Number of students	Marks	Number of students
25	2	45	4	62	2	85	2
28	2	48	3	67	5		
32	4	54	3	74	5		
36	3	55	4	78	3		
39	2	58	2	83	4		

The representation as above is known as frequency distribution. Marks are called variable (x) and the number of students against the marks is known as the frequency (f) of the variable; e.g. frequency of 67 is 5 as there are five students getting 67 marks. This representation, though better than an array, does not condense the data much and would be quite cumbersome if the number of observation increases (say from 50 to 500). A better way of condensation is to divide the observed range of variables into a suitable number of

categories or classes and to record the number of observations in each class. The above data may be presented as follows:

Table 1.3: Grouped frequency distribution of marks

Marks (x)	No. of students (f)	Marks (x)	No. of students (f)
21-30	4	61-70	7
31-40	9	71-80	8
41-50	7	81-90	6
51-60	9	Total	50 (=$\sum f_i$)

Such a table showing the distribution of the frequencies in different classes is called a frequency table and the manner in which the class frequencies are distributed over the class intervals is called grouped frequency distribution of the variable. While determining the classes in a frequency distribution table it should remembered that the classes should be clear, exhaustive, non-overlapping and preferably of equal width and the number of classes should not be too large or too small. But, if we deal with a continuous variable, it is not possible to arrange the data in the class intervals of above type. Let us consider the distribution of age in groups. If we consider the classes 0-5, 6-10, 11-15 and so on, then the persons with ages between 5 and 6 or between 10 and 11 are not taken into consideration. In such cases the classes should be taken as 0-5, 5-10, 10-15 and so on so that the upper class boundary of one class coincides with the lower class boundary of the next class.

Cumulative frequency distribution: Consider Table 1.3. Suppose one is interested to know how many students obtained marks less than or equal to 50 or how many students got more than 60. To be able to answer such questions one has to calculate the cumulative totals of the frequencies, which may be of two types-cumulative frequencies of less than type and cumulative frequencies of more than type. A table giving cumulative totals of the frequencies proceeding from the lowest class upwards is called cumulative frequencies of less than type and a table giving cumulative totals of the frequencies proceeding from the highest class of the table downwards is called cumulative frequencies of more than type. Let us now calculate these two types of cumulative frequencies (c.f.) for the above frequency distribution (i.e. for Table 1.3).

Table 1.4 Cumulative frequency (cf) distribution of marks.

Marks (Class)	No. of students	c.f. of less than type	c.f. of more than type
21-30	4	4	50 (=N)
31-40	9	13	46
41-50	7	20	37
51-60	9	29	30
61-70	7	36	21
71-80	8	44	14
81-90	6	50 (=N)	6

Thus from the above table, cumulative frequency of less than type 29 for the class 51-60 means that there are 29 students securing marks less than or equal to 60.

Similarly, cumulative frequency of more than type 37 for the class 41-50 signifies that there are 37 students who have secured more than or equal to 41.

Measures of central tendency:

According to Professor Bowley, averages are 'statistical constants which enable us to comprehend in a single effort the significance of the whole'. They give us an idea about the concentration of the values in the central part of the distribution; the value around which the series clusters. In plain terms an average of a statistical series is the value of the variable which is representative of the whole distribution. The following are the five measures of central tendency which are most commonly used: Arithmetic mean (or simply mean), Median, Mode, Geometric mean and Harmonic mean.

Arithmetic Mean (AM): Arithmetic mean of a set of observations is their sum divided by the number of observations, e.g. the arithmetic mean $\bar{x}$ of n observations $x_1, x_2, \ldots x_n$ is given by

$$\bar{x} = (x_1 + x_2 + \ldots + x_n)/n = \sum_i x_i / n$$

In case of frequency data the arithmetic mean is given by

$\bar{x} = \sum x_i * f_i / N$, where $N = \sum f_i$ and frequency of x_i is f_i .

Example 1.2: Suppose the runs scored by a cricketer in his last 9 innings are 37, 78, 25, 50, 1, 40, 15, 103 and 38.

Then $\bar{x} = (37+78+25+50+1+40+15+103+38)/9 = 387/9 = 43$.

Example 1.3: Suppose we have the following frequency distribution:

x:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

Table 1.5 Table for the calculation of $\bar{x}$

Value of variable (x)	Frequency (f)	x*f
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
Total	73 (=N)	299 ($=\sum x_i * f_i$)

Thus $\bar{x} = 299/73 = 4.09$.

The main advantages of AM are that it is easy to understand and easy to calculate; it is based on all observations and is affected least by the sampling fluctuations. The main disadvantage of AM is that it gets badly affected by extreme values and it can not be calculated if the extreme class is open.

Median: If the given values of x are arranged in an increasing or decreasing order of magnitude, then the middle-most value in this arrangement is called the median of x. Alternatively, it may be defined as a value of x such that half of the given values of x are smaller than or equal to it and half are greater than or equal to it. Thus in the above Example 1.2, to determine the median, the values (runs) are first arranged in an ascending order i.e. as 1, 15, 25, 37, 38, 40, 50, 78, 103. The median is the fifth value in this arrangement that is 38. In case of a grouped frequency distribution the median is calculated by linear interpolation using the following formula:

2.1.24. Median= $M_e = l_1 + (l_2 - l_1)*(m-c)/f$, where l_1 and l_2 are respectively the lower and the upper limit of the class in which the median lies, f=frequency of the class in which the median lies, m= N/2, and c= cumulative frequency of the class preceding the median class.

Example 1.4 Calculate the median wage of the following distribution:

Wages (in Rs.):	20-30	30-40	40-50	50-60	60-70
No. of laboueres:	3	5	20	10	5

Table 1.6 Table for calculation of median.

Wages (in Rs.)	No. of labourers (f)	Cumulative frequency (less than type)
20-30	3	3
30-40	5	8
40-50	20	28
50-60	10	38
60-70	5	43

Here, N=43, m=N/2=21.5. Cumulative frequency just greater than 21.5 is 28 and the corresponding class is 40-50. Thus the median class is 40-50. Hence l_1 =40, l_2 =50, f=20 and c=8.

So Median= $M_e = 40 + (50-40)*(21.5-8)/20 = 46.75$. Thus the median wage is Rs. 46.75

The main advantages of median are that it is easy to calculate and understand; it does not get affected by extreme values and it can be calculated for distribution with open-end class. The main disadvantage of median is that it is not based on all observations and is affected much by sampling fluctuation.

Mode: The mode of a variable is the value of the variable having the highest frequency. Thus in the Example 1.3, the value 4 occurs with the highest frequency (i.e. 17) and hence the mode is 4.

In case of grouped frequency distribution mode is determined from the following formula:

Mode= $M_o = l_1 + (f_1 - f_0)*i/[(f_1 - f_0) + (f_1 - f_2)]$, where l_1 = lower value of the modal class, i= frequency of the modal class, f_1 = frequency of the modal class, f_0 = frequency of the class preceding the modal class, f_2 = frequency of the class succeeding the modal class.

Let us now calculate the mode for the distribution given in Example 1.4. From the Table 1.6 it is clear that the maximum frequency is 20 and hence 40-50 is the modal class. Thus, $l_1 = 40$, $f_1 = 20$, $f_0 = 5$, $f_2 = 10$ and $i = 10$. So, Mode= $M_o = 40+(20-5)*10/[(20-5)+(20-10)] = 46$.

The main advantage of mode is that it is readily comprehensible and easy to calculate and that it does not get affected by extreme values. The main disadvantage of mode is that it is not always possible to find a clearly defined mode and also it is not based on all observations.

Geometric Mean (GM): If a variable has n values $x_1, x_2, \ldots x_n$, then its geometric mean x_g is defined by $x_g = (\Pi\, x_i)^{1/n}$. It may be observed that $\log x_g = \sum \log x_i / n$. Thus the logarithm of the GM of a variable is the AM of its logarithms. Thus the geometric mean of four numbers 2, 3, 5 and 10 is $x_g = (2*3*5*10)^{1/4} = 4.16$. It may also be computed in the following way.

x	log x
2	0.69
3	1.10
5	1.61
10	2.30
Total	5.70

$\log x_g = 5.70/4 = 1.43$ i.e. $x_g = \exp(1.43) = 4.16$.

Geometric mean is commonly used to find the rate of population growth and the rate of interest.

Harmonic Mean (HM): Harmonic mean of a variable x with given values $x_1, x_2, \ldots x_n$ is defined by

$x_h = n/(\sum 1/x_i)$ or $1/x_h = 1/n * \sum 1/x_i$. For the numbers 2.3, 5 and 10, the HM is

$x_h = 4/(1/2+1/3+1/5+1/10) = 4/1.13 = 3.54$.

Measures of dispersion:

The measures of central tendency give a general idea as to the whole set of values. They, however, do not tell us the degree of variability present in the series. Measures of dispersion aim at measuring this degree of scatter or variability about the average. Mainly three different measures are used to determine this feature of a variable which is called the scatter or dispersion. They are Range, Mean deviation and Standard deviation.

Value (run) x_i	x_i -43	$(x_i-43)^2$
37	-6	36
78	35	1225
25	-18	324

2.1.33. Range: The simplest measure of the dispersion of a variable is its range, which is defined as the difference between the highest and the lowest given values. Thus in the Example 1.1, the range is 85-25=60. The range is the easiest to compute among all measures of dispersion and most easily comprehensible. It however, uses only two extreme observations of the series and not all the observations which makes it an inferior measure of dispersion.

Mean deviation: If A be the chosen average value of the variable x, then $x_i - A$ is the deviation of the i-th given value of x from the average (i=1,2,….. n). Then the Mean deviation about A is defined as $MD_A = \sum | x_i - A| / n$ for ungrouped data and $MD_A = \sum | x_i - A|^* f_i / N$ for grouped data; where $N= \sum f_i$ and $| x_i - A|$ is the absolute value of $(x_i - A)$. It can be shown that the mean deviation is least when measured about the median.

Let us consider the Example 1.2 and calculate the mean deviation about mean for the set of given data. Here $A= \bar{x}=43$. We construct the following table to carry out the necessary calculations.

The mean deviation about mean is thus 204/9= 22.67 where as the range of the same series is 103-1=102.

Value (run) x_i	x_i -43	$\|x_i-43\|$
37	-6	6
78	35	35
25	-18	18
50	7	7
1	-42	42
40	-3	3
15	-28	28
103	60	60
38	-5	5
Total	0	204

Although mean deviation takes into account all the observations it is computationally not as easy as range is. Moreover, it is not easily amenable to algebraic treatment.

Standard Deviation (SD): The standard deviation of a set of n variables $x_1, x_2, \ldots x_n$ is computed as

$\sigma_x = \sqrt{(\sum (x_i - \bar{x})^2 / n)}$ for ungrouped data and $\sigma_x = \sqrt{(\sum (x_i - \bar{x})^2 * f_i / N)}$ for grouped data where $N= \sum f_i$. Let us consider the Example 1.2 again and calculate the S.D. We construct the flowing table for computation of the S.D. The square of SD is called the variance and is denoted by σ_x^2.

50	7	49
1	-42	1764
40	-3	9
15	-28	784
103	60	3600
38	-5	25
Total	0	7816

Here $\bar{x}=43$.

$\sum (x_i - \bar{x})^2 = 7816$; $n=9$

$\sigma_x = \sqrt{7816/9} = 29.47$.

Var of X= $\sigma_x^2 = 7816/9 = 868.4$

The standard deviation involves a little more computational hazard than range; but it takes into consideration all the values of the series and is readily amenable to algebraic treatment. As such, SD is generally regarded as the best measure of dispersion.

Quartile Deviation: The quartiles are points that divide the frequency distribution of the variable into four equal parts. Thus the first quartile Q_1 is the value of the variable such that 25% of the total number of observations are less than or equal to it. Similarly, the third quartile Q_3 is the value of the variable such that 75% of the total number of observations are less than or equal to it and second quartile Q_2 is the median. The quartile deviation is measured as $Q = (Q_1 + Q_3)/2$.

Measure of relative distribution: The measures of dispersions we have discussed so far are all expressed in the same units as those of the variable. As such, they can not be used in comparing two distributions of different types with respect to their variability. For purposes of such comparison, a measure of dispersion has to be made unit free. The simplest procedure is to express a measure of dispersion as a percentage of a measure of central tendency. The most commonly used measure of relative dispersion is the coefficient of variation C.V. = $100*\sigma_x / \bar{x}$ where $\bar{x}$ is supposed to be non-zero. For comparing two series we calculate the coefficient of variations for each series. The series having greater C.V. is said to be more variable than the other and the series having lesser C.V. is said to be more consistent. The coefficient of variation also gives a true picture of relative accuracy of the two series. Thus for the above example, when $\sigma_x=29.47$, $\bar{x}=43$, C.V. = 29.47*100/43=68.53.

Example 1.5 An analysis of monthly wages paid to the workers in two firms A and B belonging to the same industry gives the following results:

	Firm A	Firm B
Number of workers	500	600
Average monthly wages	Rs. 186	Rs. 175
Variance of distribution of wages	81	100

Determine in which firm there is greater variability in individual wages?

C.V. of distribution of wages for firm A = $100*\sqrt{81}/186 = 100*9/186 = 4.84$
C.V. of distribution of wages for firm B= $100*\sqrt{100}/175 = 100*10/175 = 5.71$
As the C.V. for firm B is greater than C.V. for firm A, firm B has greater variability in individual wages.

Correlation and Regression

In real life we may come across certain series where each item of the series may assume the values of two or more variables. For example, if we measure the heights and weights of a certain group of persons, we shall get what is known as bivariate distribution- one variable relating to height and the other to weight. In a bivariate distribution we may be interested to find out if there is any relation between the two variables. Correlation is a measure of association between such variables. It measures how strongly the variables are related, or change, with each other. If two variables tend to move up or down together, they are said to be positively correlated. If they tend to move in opposite directions, they are said to be negatively correlated. More specifically correlation *is a measure of the degree of linear relationship between two variables*, usually labeled X and Y. Thus X may be the height of a group of individuals and Y their weight. Regression is a statistical technique used to establish the relationship of a dependent variable (e.g. excess return) and one or more independent variables (e.g. exposure to market, size, and value risks). While in regression the emphasis is on predicting one variable from the other, in correlation the emphasis is on the degree to which a linear model may describe the relationship between two variables. In regression the interest is directional, one variable is predicted and the other is the predictor; in correlation the interest is non-directional, the relationship is the critical aspect.

The simplest way of diagrammatic representation of bivariate data is the use of a scatter diagram. Taking two perpendicular axes of coordinates, one for x and the other for y, each pair of values is plotted as a point on graph paper. The whole set of points taken together constitutes the scatter diagram. If it is found that as one variable increases (decreases), the other also increases (decreases) in general or on the average, then there exists a positive correlation between them (fig. 2.1). On the other hand, a negative correlation exists if one variable increases (decreases) with the decrease (increase) of the other (fig. 2.2). There may be a third situation where as one variable increases (decreases), the other remains constant. This is the case of zero or no correlation and the variables are said to be uncorrelated.

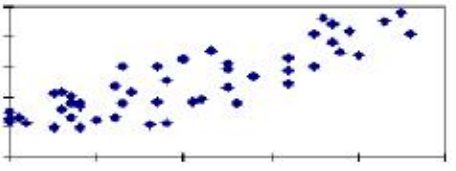


Fig. 2.1

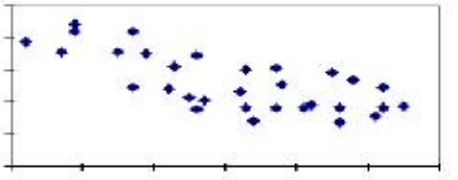


Fig. 2.2

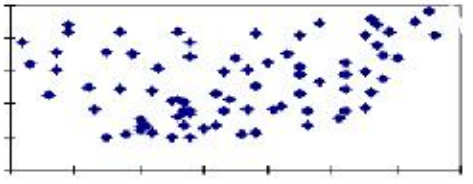


Fig.2.3

Correlation Coefficient: Suppose we have n pairs of values (x_1, y_1), (x_2, y_2),.....(x_n, y_n). Then the *Pearson Product-Moment Correlation Coefficient* (r) is defined as

$r(x,y) = Cov(x,y)/(SD(x)* SD(y))$, where

SD (x) = Standard Deviation of x $=\sqrt{(\sum(x_i - \bar{x})^2/n)} = \sqrt{(\sum x_i^2/n - \bar{x}^2)}$
SD (y) = Standard Deviation of y $=\sqrt{(\sum(y_i - \bar{y})^2/n)} = \sqrt{(\sum y_i^2/n - \bar{y}^2)}$
Cov (x,y)= Covariance (x,y)= $\sum(x_i - \bar{x})^* (y_i - \bar{y})/n. = \sum x_i * y_i/n - \bar{x} * \bar{y}$

Example 2.1 Calculate the correlation coefficient for the following heights in inches of fathers (x) and their sons (y).

X:	65	66	67	67	68	69	70	72
Y:	67	68	65	68	72	72	69	71

Table for calculation of correlation coefficient

X	y	x^2	y^2	x*y
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
$\sum x$=544	$\sum y$=552	$\sum x^2$=37028	$\sum y^2$=38132	$\sum xy$=37560

$\bar{x} = \sum x/n = 544/8=68$; $\bar{y} = 552/8 =69$; SD (x)= $\sqrt{(37028/8 - (68)^2)} = \sqrt{4.5}=2.12$;
SD (y)= $\sqrt{(38132/8 - (69)^2)} = \sqrt{5.5}=2.35$; Cov (x,y)=37560/8- 68*69 =3
r (x,y)=3/(2.12*2.35)=0.6 i.e. the correlation coefficient between x and y is 0.6

Short-cut method

X	y	$u=x-\bar{x}$ i.e. u= x-68	$v=y-\bar{y}$ i.e. v=y-69	u^2	v^2	u*v
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
67	68	-1	-1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
Total		0	0	36	44	24

$\bar{u} = 0$, $\bar{v} = 0$; SD (u) = $\sqrt{36/8}=\sqrt{4.5}=2.12$, SD (v)= $\sqrt{44/8}= \sqrt{5.5}=2.35$;
Cov (u,v) = 24/8=3; r(u,v)= 3/(2.12*2.35)=0.6 =r (x,y).

Properties of correlation coefficient: (i) The correlation coefficient of x and y is a pure number, that is, it is independent of the units of measurement of x and y.
(ii)The value of r lies between –1 and +1 i.e. $-1 \le r \le 1$.

Limitation of correlation coefficient: (i)The main drawback of the correlation coefficient is that it measures the degree of linear relationship between two variables. Thus if x and y have a non-linear relationship (e.g. $y=x^2$), then r will be very small (or zero), although actually the relationship between the variables may be quite strong. It is, therefore, advisable that one should see whether the general relationship between x and y is linear before using r as a measure of this relationship.
(ii) Another serious drawback of correlation is that 'two variables are correlated' does not necessarily mean that they are causally related-that one variable is the cause of the other. Indeed, the two variables may appear to be correlated even when both are caused by some other variable or variables. For example, one may find a strong negative correlation between the height and marks obtained by a group of students in a particular examination although in reality none may be said to be caused by the other. Such types of correlation often called spurious correlation may be very misleading and hence proper care should be taken in such cases.

Regression: Regression is a statistical technique used to establish the relationship of a dependent variable (e.g. excess return) and one or more independent variables (e.g. exposure to market, size, and value risks). While in regression the emphasis is on predicting one variable from the other, in correlation the emphasis is on the degree to which a linear model may describe the relationship between two variables. In regression the interest is directional, one variable is predicted and the other is the predictor; in correlation the interest is non-directional, the relationship is the critical aspect.

Lines of regression: If the variables of a bivariate distribution are related, we find that the points in the scatter diagram tend to cluster around some curve called the 'curve of regression'. If this curve is a straight line, it is called the line of regression. The line of regression is the line, which gives the best estimate to the value of one variable for any specific value of the other variable. Thus the line of regression is the line of 'best fit' and is obtained by the principles of least squares.

Let us suppose that in a bivariate distribution (xi,yi); I=1,2,...n; Y is the dependent and X is the independent variable. Let the line of regression of Y on X be $Y= a + b*X$ ———(1)

According to the principle of least squares, the normal equations for estimating a and b are
$\sum y_i = n*a + b* \sum x_i$ and $\sum x_i y_i = a * \sum x_i + b* \sum x_i^2$.
Solving these two normal equations for a and b, we get
$b= Cov(x,y)/ var(x) = r(x,y)* \sigma_y / \sigma_x$ where σ_x = SD (x); σ_y= SD (y); and
$a= \bar{y} - b* \bar{x}$

substituting these values of a and b in equation (1) above, the regression line of y on x becomes

$Y= \bar{y} + (r*\sigma_y / \sigma_x)* (X- \bar{x})$.

Where a is the y-intercept of the line and b is its slope. The coefficient b is the amount by which Y increases for a unit increment in the value of x.. It is called the regression coefficient of y on x sometimes denoted by b_{yx}.

Similarly if we are interested in predicting x from y, we use the regression line of x on y, which has the equation $X= \bar{x} + (r*\sigma_x / \sigma_y)* (Y- \bar{y})$.
$(r*\sigma_x / \sigma_y)$ which is the amount by which X increases for a unit increment in y, is the regression coefficient of x on y sometimes denoted by b_{xy}.

Let us take the Example 2.1 and obtain the equations of the lines of regression. Also obtain the estimate of Y for X=70.

The values of $\bar{x}$, $\bar{y}$, σ_x, σ_y and r have already been calculated.
$\bar{x}$=68, $\bar{y}$=69, σ_x=SD (x)=2.12, σ_y=SD(y)=2.35 and r=0.6

Equation of line of regression of Y on X is
$Y= \bar{y} + (r*\sigma_y / \sigma_x)* (X- \bar{x})$
i.e. Y=69+(0.6*2.35/2.12)*(X-68) =>Y=0.665*X+23.78
Equation of line of regression of X on Y is
$X= \bar{x} + (r*\sigma_x / \sigma_y)* (Y- \bar{y})$
i.e. X=68+(0.6*2.12/2.35)*(Y-69) => X= 0.54*Y+30.74

To estimate Y for given X we use the line of regression of Y on X. If X=70, the estimated value of Y is given by $\hat{Y}$=0.665*70+23.78=70.33.

It can be proved that correlation coefficient (r) is the geometric mean of the two regression coefficients b_{xy} and b_{yx}.

Time Series Data

Arrangement of statistical data in chronological order i.e. in accordance with occurrence of time is known as Time series. In other words, time series refers to statistical data which relate to successive intervals or points of time. Examples of time series are yearly, quarterly or monthly production or consumption figures for a particular commodity, price of a commodity at different points of time, profits earned by a company for each of the past five years, the weekly wholesale price index for the last 30 weeks etc. Symbolically, y_t denotes the value of the variable at time t (t=1, 2,.....n). In case the figures relate to n successive periods (and not points of time), t is to be taken as the mid-point of the t-th period.

The analysis of time series enables us to understand the past behaviour or performance in order to predict the future better; it also helps in business planning.

Components of time series

A time series may contain one or more of the following four components:

1. **Secular trend (T_t):** By secular trend (or simply, trend) of a time series we mean the smooth, regular, long term movement of the series if observed long enough. Thus certain series may show an upward trend e.g. population or industrial production of a country may show an upward trend over a period of time. Again certain series may exhibit a downward trend e.g. mortality rate may show a steady decline over time. Any sudden or abrupt change is incompatible with the concept of trend.

2. **Seasonal fluctuation (S_t):** By seasonal fluctuations we mean a periodic movement in the time series where the period is no longer than one year. A periodic movement in a time series is one which recurs or repeats at regular interval of time (or periods). Examples of seasonal fluctuation may be found in the passenger traffic during the 24 hours of a day, sales of a departmental store during the 12 months of a year and so on.

3. **Cyclical fluctuations (C_t):** By cyclical fluctuation we mean the oscillatory movement in a time series, the period of oscillation being more than a year. One complete period is called a cycle. The cyclical fluctuations are not necessarily periodic, as the length of the cycle as also the intensity of fluctuations may change from one cycle to another. Every businessman is familiar with the alternating periods of prosperity and depression in business which follow one another in an irregular manner.

4. **Irregular component (I_t):** These are components which are either wholly unaccountable or are caused by such unforeseen events as wars, floods, strikes etc. This category of movements includes all types of variation that are not accounted for by secular trend or seasonal or cyclical fluctuations.

Concept of moving average

The method of moving average consists in measurement of trend by smoothing out the fluctuations of the data by means of a moving average. Moving average of period m of a time series gives us a new series of successive averages (arithmetic means) each of m successive observations of the time series. Thus the first average is the mean of the first m terms, the second average is the mean of m terms from 2^{nd} to (m+1) th term, the third is the mean of the m terms starting from 3^{rd} to (m+2) th term.

If m is odd = (2K+1) say, moving average is placed against the mid-value of the time interval it covers, i.e. against t=k+1 and if m is even = 2k say, it is placed between the two middle values of the time interval it covers, i.e. between t=k and t=k+1. In the later case the moving average does not coincide with an original time period and an attempt is made to synchronize the moving averages and the original data by centering the moving averages which consists in taking a further moving average of period two of these moving averages and putting the first of these values against t=k+1. The graph obtained by plotting the moving averages against time series gives trend.

In this method, the main problem which is of paramount importance, lies in determining the period of the moving average that will completely eliminate the oscillatory movements affecting the series. It can be mathematically established that if the fluctuations are regular and periodic, then the moving average completely eliminates the oscillatory movement provided (i) the extent of moving average is exactly equal to or a multiple of the period of oscillation and (ii) the trend is linear. Since in practical situations different cycles vary in amplitude and period, in such cases the appropriate period of moving average would be equal to or somewhat greater than the mean period of the cycles in the data.

Moving average method is very flexible in the sense that the addition of a few more figures to the series simply results in some more trend values, not altering the previous calculations. One major disadvantage of this method is that it does not provide trend values for all the time points; e.g. for a moving average of period m =2k+1, we have to forego the trend values for the first k and the last k terms of the series. Also when the trend is not linear, moving average does not depict an unbiased picture.

Example 3.1 Assuming a four-yearly cycle, calculate the trend by the method of moving averages from the following data relating to production of a commodity 'X'.

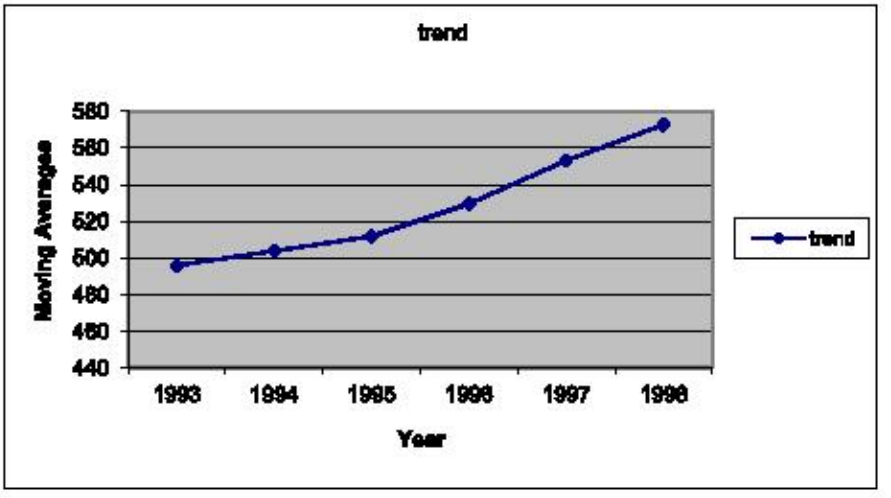
Year	Production ('000 Tonnes)	Year	Production('000Tonnes)
1991	464	1996	540
1992	515	1997	557
1993	518	1998	571
1994	467	1999	586
1995	502	2000	612

Here since the period of the cycle is four, we calculate a moving average of period four, which being an even number needs a further two point moving average to be calculated for centering. The calculations are shown in the table given in the next page.

Year	Production (000 tons)	4-yearly moving total	4-yearly moving average	4-yearly moving average-centered
1991	464			
1992	515	1964	491	
1993	518	2002	500.5	495.7
1994	467			503.6

		2027	506.75	
1995	502			511.6
		2066	516.5	
1996	540			529.5
		2170	542.5	
1997	557			553
		2254	563.5	
1998	571			572.5
		2326	581.5	
1999	586			
2000	612			

The following graph obtained by plotting the moving averages against time series gives the trend.



Growth rate calculation with time series data

Calculation of growth rate from time-series data is of great importance to a wide spectrum of people in the society including statisticians, economic planners, market researchers, businessmen, academicians and politicians. Two types of growth rates are most commonly observed in time series data viz. linear growth and exponential growth.

If we believe that a series y_t will increase in constant absolute amounts in each time period, then the growth is said to be linear. Thus for a linear growth rate, the growth curve follows the equation $y_t = a + b*t$, where t is the time and y_t is the value of y at time t (t=1,2,…n). The constants a, and b are obtained by solving the normal equations

$\sum y_t = n*a + b* \sum t$ and
$\sum t*y_t = a* \sum t + b* \sum t^2$, t=1,2,….n.

On the other hand, if we assume that the series y_t grows with constant ratio rather than constant absolute amount, then the growth is said to be exponential. An exponential growth

curve is to be used when we are interested more in the relative changes in the variable rather than the absolute magnitudes of change. In that case, y_t follows an exponential growth curve $y_t = a*e^{r*t}$, where $a= y_0$, the value of y_t at time 0 (i.e. in the initial year), t is the time and y_t is the value of y at time t (t=1,2,...n) and r is the growth rate. Examples of exponential growth include bank accounts on which fixed interest is accumulating, human population unhindered by predation or environmental problems, and contagious diseases for which no immunization is available. The constants of equations can be obtained by solving the normal equations

$\sum \ln (y_t) = n*\ln (a) + r* \sum t$ and
$\sum t* \ln (y_t) = \ln (a)* \sum t + r* \sum t^2$, t=1,2,....n.

Thus in simple term, linear growth means that the original value increases periodically by a set amount, where as exponential growth means that the original value increases periodically by a set percentage.

Example of linear growth: Suppose it is given that a baby grows by 0.25 inches per month till she is four years old. Determine how long (tall) will she be at the end of 10^{th} month if she starts with 20 inches, and when will she be 30 inches tall.
Here $a= y_0 = 20$ inches, b= 0.25 inches per month.
Hence the length of the baby at the end of 10^{th} month will be $y_{10} = 20+0.25*10 = 22.5$ inches. For the baby to be 30 inches tall we have to determine 't' for which $y_t = 30$; i.e. to find t when a=20, b=0.25 and $y_t = 30$. Using the linear equation, 30=20+0.25*t or t=40 months. Thus the baby will be 30 inches tall after 40 months from the starting month.

Example of exponential growth: Suppose a study shows that the population of a particular region (town) has been growing at some constant rate. The population of the town was 400,000 in the year 1981 and 600,000 in the year 1991. Determine the rate at which the population of the town grew over this period?

Here yt = 600,000, y_0 = 400,000 and t=10. If r denotes the growth rate, then

$600000 = 4000000* e^{r*10} \Rightarrow e^{r*10} = 1.5 \Rightarrow r = \ln (1.5)/10 = 0.04$ which means that the rate of growth of the population is 4% per annum.

Probability and Simple Random Sampling

The word probability may be used in two different contexts. First it may be used in regard to some proposition. Take for instance, the statement 'It is very probable that the country's GDP will grow at 7% in the next two years'. Probability here means the degree of belief in the proposition of the person making the statement. This may be called subjective probability.

Alternatively, the word may be used in regard to the results of an experiment that can be repeated under essentially similar conditions. The results of such experiments are called *events*. The probability of an event here refers to the proportion of cases in which the event occurs in such repetitions of the experiment. This type of probability may be called objective. It is generally in this sense that the term 'probability' is used in Statistics.

Definition of various terms/concepts:

Random experiment: An experiment the outcome of which depends on chance. For example throwing a die or tossing a coin.

Events: Any outcome/result of an experiment is called an event; e.g. getting '1' from a single throw of a die or getting a 'Head' from a single toss of a coin.

Elementary events: The results of an experiment that can not be decomposed further are called elementary events and the whole set of elementary events is called sample space. Thus getting '1' from a single throw of a die or getting a 'Head' from a single toss of a coin are elementary events.

Compound events: Events obtained by combining together two or more elementary events are called compound events. Thus occurrence of an even number of points is a compound event consisting of 3 elementary events viz. getting '2' or '4' or '6' from a throw of a die.

Mutually Exclusive events: Events are said to be called mutually exclusive if the happening of any one of them precludes the happening of all others, i.e. if no two or more of them can happen simultaneously in the same trial. In throwing a die all the six\faces numbered 1 to 6 are mutually exclusive since if any one of these faces comes, the possibility of others, in the same trial is ruled out.

Equally Likely events: Outcomes of a trial are said to be equally likely if taking into consideration all the relevant evidences, there is no reason to expect one in preference to other. In tossing an unbiased or uniform coin, getting a head or a tail are equally likely events.

Favourable cases: The number of cases favourable to an event is the number of outcomes which entail the happening of the event. Thus in drawing a card from a pack of cards the number of cases favourable to drawing an 'ace' is 4, for drawing a 'spade' is 13 and for drawing a 'red card' is 26.

Union of events: The union of two events A and B, denoted by $(A \cup B)$ implies the occurrence of either A or B or both A and B i.e. it denotes the occurrence of at least one event out of A and B.

Intersection of events: Intersection of two events A and B denoted by $(A \cap B)$ means the occurrence of both the events A and B i.e. it denotes simultaneous occurrence of both the events A and B.

Complement of an event: Complement of an event A denoted by A^c denotes the non-occurrence of A.

Exhaustive events: If the union of a set of events contains all possible elementary events then that set of events is known as the exhaustive set of events. Thus the events 'getting a head' and a 'getting a tail' in a single toss of a coin are exhaustive events.

Independent events: Several events are said to be independent if the occurrence or non-occurrence of an event is not affected by the occurrence or non-occurrence of other events.

Classical definition of probability: If an experiment results in n exhaustive, mutually exclusive and equally likely elementary events and if m of them are favourable to a particular event A, then the probability of happening of A is given by $P(A) = m/n$ $(m \leq n)$. Thus if we consider the experiment of throwing an unbiased die, then there are in total six elementary events which are exhaustive, mutually exclusive and equally likely and they are {1,2,3,4,5,6}. Thus n=6. Now if we define the event A as getting an even number from a single throw of a die then the number of cases favourable to A is 3 and they are {2,4,6}, i.e. m=3. So P(A)=3/6=0.5. Since for a sure event m=n, Probability of a sure event is 1. Also, as for an impossible event m=0, Probability of an impossible event is zero.

Addition theorem of probability: If A and B are two events associated with a random experiment, then $P(A \cup B)=P(A)+P(B)-P(A \cap B)$. If A and B are mutually exclusive, i.e. if $P(A \cap B)=0$, then $P(A \cup B)=P(A)+P(B)$. Also since by definition A and A^c are exhaustive and mutually exclusive, and $(A \cup A^c)$ is a certain or sure event, $P(A) + P(A^c)=1$ or $P(A^c)=1-P(A)$.

Conditional probability: If A and B are two events associated with a random experiment, with $P(B) \neq 0$, then the probability of occurrence of A given that the event B has already occurred is called the conditional probability of A given B and it is denoted by P(A/B). $P(A/B)=P(A \cap B)/P(B)$.

Statistical independence of events: Two events are said to be statistically independent if the occurrence or non-occurrence of an event does not affect the probability of the occurrence or non-occurrence of the other event. That is in case of independent events P(A/B)=P(A) and P(B/A)=P(B).

Multiplication law of probability: For two events A and B,
$P(A \cap B)=P(A)*P(B/A)$, $P(A)>0$ or $P(A \cap B)=P(B)*P(A/B)$, $P(B)>0$.

In case of independent events since P(B/A)=P(B) and P(A/B)=P(A),
$P(A \cap B)=P(A)*P(B)$

Random variable: It may be seen that outcome of an experiment need not be number, for example, the outcome of a single toss of a coin can be 'head' or 'tail'. However, we often

want to represent outcomes as numbers. A random variable is a function that associates a unique numerical value with every outcome of an experiment i.e. it is a real valued function defined on the sample space of a random experiment. Mathematically, it is described by its probability distribution, which specifies the possible values of the random variable together with the probability associated with each value. For example, we may define the random variable X as the number obtained from a single throw of an unbiased die. Then X may take any of the six values {1, 2, 3, 4, 5, 6} each having the probability 1/6. Similarly, we may define another random variable Y as the number of heads obtained in two toss of a perfect coin. Then Y can take any one of the three values 0, 1 or 2 with probabilities 1/4, 1/2 and 1/4 respectively.

Expected Value: The expected value of a random variable indicates its average or central value. It is a useful summary value of the variable's distribution. The expected value of a random variable is symbolized by E(X) or μ.

If X is a random variable with possible values $x_1, x_2, \ldots x_n$ and $p(x_i)$ denotes $P[X=x_i]$, the the expected value is defined as $E(X)= \sum x_i * p(x_i)$, where the elements are summed over all values of the random variable X.

Example 2.1 Define a random variable X= the number obtained from a single throw of an unbiased die,
$E(X) = (1*1/6) + (2*1/6) + (3*1/6) + (4*1/6) + (5*1/6) + (6*1/6) = 3.5$.

Example 2.2 Define a random variable Y= number of heads obtained from two tosses of a perfect coin. Then the probability distribution of Y will be as under.

Value of Y:	0	1	2
Probability:	1/4	1/2	1/4

$E(Y)= (0*1/4) + (1*1/2) + (2*1/4) = 1$

Variance: The variance of a random variable is a non-negative number which gives an idea of how widely spread the values of the variable are likely to be; the larger the variance, the more scattered the observations on average. It is a measure of the spread of a distribution about its average value. Variance is symbolized as V(X) or Var (X) or σ^2. The variance of a random variable X is defined to be $V(X) = E[X-E(X)]^2 = E(X^2) - \{E(X)\}^2$, where E(X) is the expected value of the random variable X

Let us calculate the variance for Example 2.1.

Value of X:	1	2	3	4	5	6
Probability:	1/6	1/6	1/6	1/6	1/6	1/6

$E(X^2)= (1^2*1/6)+(2^2*1/6)+(3^2*1/6)+(4^2*1/6)+(5^2*1/6)+(6^2*1/6)$
$=(1*1/6)+(4*1/6)+(9*1/6)+(16*1/6)+(25*1/6)+(36*1/6)=91/6=15.17;$
Also $E(X) =3.5$
Therefore, $V(X) =15.17-12.25=2.92$

Similarly, in Example 2.2, $E(Y^2)= (0^2*1/4)+(1^2*1/2)+(2^2*1/4)=(1*0.5)+(4*0.25)=1.5$ and $E(Y)=1$.
Hence $V(Y) =1.5-1=0.5$

Some elementary concepts and definitions of sample survey:

Unit: An elementary unit or simply a unit is an element or a group of elements, on which observations can be made or from which the required statistical information can be ascertained. The type of unit to be considered depends on the purpose of survey; e.g. a family or household may be considered as the unit in a family budget enquiry, whereas a unit may be a farm or plot in a crop survey or a factory may be taken as a unit in an industrial survey. A reporting unit is the unit, which actually supplies the required statistical information, or from which the information can be conveniently ascertained and such a unit may be elementary unit itself or a unit representing a group of elementary units. For instance the head of the family may be the reporting unit supplying the information on individual members of the family.

Population: The collection of all units of a specified type in a given region at a particular point or period of time is termed a population or universe. Thus we may consider a population of persons, families, farms, cattle, houses or automobiles in a region or a population of fish in a tank etc. depending on the nature of data required. A population is said to be finite or infinite according as the number of units in the population (called population size) is finite or infinite. Therefore the population must be capable of division into what are called *sampling units* for purposes sample selection. The sampling units must cover the entire population and they must be distinct, unambiguous and non-overlapping in the sense that every element of the population belongs to one and only one sampling unit.

Population Parameter: Any function of the values of all the population units is known as population parameter. Suppose a finite population consists of the N units $U_1, U_2, \ldots U_N$ and let Y_i be the value of the variable y, the characteristics under study, for the i-th unit U_i (i=1,2,….N). For instance, the unit may be a household and the variable y may be the total consumer expenditure during a particular month. A population parameter may be the total or average consumer expenditure of all the households in the region. Some of the important parameters usually required to be estimated in surveys are Population Total, Population Mean and Population Variance which are defined below:

Population Total= $Y = Y_1 + Y_2 + \ldots + Y_N = \sum Y_i$ = Sum of all Y_i s, i=1,2…..N
Population Mean= $\bar{Y} = Y/N = \sum Y_i/N$ =Arithmetic Mean of Y_i s.
Population Variance= $\sigma^2 = \sum(Y_i - \bar{Y})^2 / N$ = Variance of Y_i s. σ^2 is a measure of inherent heterogeneity in the population

Sample and statistic: A sample is a collection of units selected from a population according to some specified procedure. If a sample is drawn in such a manner that each unit in the population has a predetermined probability of being selected then it is called a random sample. The sampling units selected in the sample are called sample units and the values of the characteristics under study for the sample units are called sample observations. The number of sampling units selected in a sample is called sample size and the ratio of sample size to the total number of population units is termed sampling fraction.
Suppose a sample of size n units is selected from a population of N units according to some scheme and let the sample observations be denoted by $y_1, y_2, \ldots y_n$. Any function of these values, which is free from the unknown population parameter, is called a Statistic.
For example, the sample mean $\bar{y} = (y_1 + y_2 + \ldots + y_n)/n$ is a statistic.

Estimator: An estimator is a statistic obtained by a specified procedure for estimating a population parameter. The particular value, which the estimator takes for a given sample is known as estimate. An estimator T is said to be an unbiased estimator of the population parameter θ, if $E(T) = \theta$.

Sampling/Sample Survey: Sampling means selecting a part of the population (called sample) to represent the whole population. In sample survey, the units selected in the sample are surveyed and an inference is drawn about the population on the basis of these observations.

Probability Sampling: If all the population units have a predetermined probability of being selected in the sample, then the sampling is called probability sampling.

Simple Random Sampling (SRS): In simple random sampling, each member of a population has an equal chance of being included in the sample. Also, each combination of equal number of members of the population has an equal chance of composing the sample. These two properties are what define simple random sampling. To select a simple random sample, one needs to list all of the units in the survey population. Simple random sampling can be done with replacement (SRSWR) or without replacement (SRSWOR). In SRSWR, sample units are drawn from the population one by one, after each drawing the individual unit selected being returned to the population. On the other hand, in SRSWOR, sample units are drawn one by one but the unit obtained at any drawing is not returned to the population and the next draw is made from among the remaining units. Thus a sample with replacement means that there is a possibility that the sampled units may be selected twice or more while in simple random sampling without replacement a unit once selected can not be selected again.

Simple random sampling (SRS) is the simplest of all probability sampling. Since all the population units get equal chances of being selected in the sample in SRS, the element of subjectivity or personal bias is completely eliminated. One of the serious limitations of SRS is that it requires an up-to-date frame, i.e. a completely categorized population from which samples are to be drawn. Very often it becomes virtually impossible to identify the units in the population before the sample is drawn. Again a simple random sample may result in the selection of the sampling units which are widely spread geographically and in such cases the cost of collecting the data is much both in terms of money and time. Also for a given precision, simple random sampling usually requires larger sample size as compared to stratified random sampling.

Estimation of population mean: Suppose from a population of size N, a sample of size n is drawn by SRSWR and also by SRSWOR. Let Y be the character under study and let the sample observations be denoted by $y_1, y_2, \ldots y_n$.
Then for both SRSWR and SRSWOR, the sample mean $\bar{y} = (y_1 + y_2 + \ldots + y_n)/n$ gives an unbiased estimate of the population mean.
An unbiased estimate of the population total is given by $N^* \bar{y}$.

Example: Draw a random sample of size 10 from a population of size 400 by (i) SRSWR and also by (ii) SRSWOR by making use of the random numbers given below.

2952	6641	3992	9792	7969	5911	3170	5624
4167	9524	1545	1396	7203	5356	1300	2693

Soln. First of all we identify the 400 units in the population with the numbers 1 to 400 as U_1, U_2, ... U_{400}. Then we shall draw three-digit random numbers from the given set of random numbers starting from first row and first column of the series i.e. from 2 and then move row-wise. To ensure equal probability for each individual, we shall take the numbers 001 to 800 (the greatest three digit multiple of 400) and shall ignore all other 3-digit numbers. We shall divide the random number drawn by 400 and take the remainder. The remainder here may vary from 000 to 399. The remainders 001 to 399 will correspond to the unit of the same numbers, where as the remainder 000 will correspond to the 400^{th} unit.

Random number drawn	Remainder when divided by 400	Serial number of the unit selected for SRSWR	Serial number of the unit selected for SRSWOR
295	295	295	295
266	266	266	266
413	013	13	13
992	Rejected	-	-
979	Rejected	-	-
279	279	279	279
695	295	295	Rejected as it has already been included
911	Rejected	-	-
317	317	317	317
056	056	56	56
244	244	244	244
167	167	167	167
952	Rejected	-	-
415	015	15	15
451	051	Sample of size 10 has already been selected	51

In SRSWR, the units of the population selected in the sample of size 10 are those with numbers 295, 266, 13, 279, 295, 317, 56, 244, 167 and 15.

In SRSWOR, the units of the population selected in the sample of size 10 are those with numbers 295, 266, 13, 279, 317, 56, 244, 167, 15 and 51.

Probability Proportional To Size Sampling, Stratified Sampling

Under certain circumstances, selection of units with unequal probabilities provides more efficient estimators than equal probability sampling, and this type of sampling is known as unequal or varying probability sampling. In the most commonly used varying probability sampling scheme, the units are selected with probability proportional to some measure of their size (PPS) where the size measure is the value of an auxiliary variable X related to the characteristic Y under study and this sampling scheme is termed probability proportional to size sampling. For instance, the number of persons in some previous period may be taken as a measure of the size in sampling area units for a survey of socio-economic characters, which are likely to be related to population. Similarly, in estimating crop characteristics the geographical area or cultivated area for a previous period, if available, may be considered as a measure of size, or in an industrial survey, the number of workers may be taken as the size of an industrial establishment.

Since a large unit, that is, a unit with a large value for the study variable Y, contributes more to the population total than smaller units, it is natural to expect that a scheme of selection which gives more chance of inclusion in a sample to larger units than to smaller units would provide estimators more efficient than equal probability sampling. Such a scheme is provided by PPS sampling. This technique has found wide application in multi-stage sampling where the first stage units (say, villages) may differ considerably in size.

Practical procedure of PPS sampling: The procedure of sampling with PPS essentially consists in associating each unit a number of numbers equal to or exactly proportional to its size and selecting the unit corresponding to a number chosen at random from the totality of numbers so associated. Two methods which are used most often for drawing a PPS sample are (i) Cumulative Total Method and (ii) Lahiri's Method.

(i) Cumulative Total Method: Let the size of the i-th population unit U_i be X_i, i=1,2,...N. A straight forward application of the principle of associating with each unit numbers equal to its size is to associate the numbers 1 to X_1 with the first unit, the numbers X_1+1 to $X_1 + X_2$ with the second unit and so on. The total number of numbers so associated is X, the sum of the sizes of all the units in the population. Then a number R is chosen at random from 1 to X and the unit with which this number is associated is considered as selected. This procedure of selection is termed cumulative total method, since this method needs cumulating of the sizes.

The steps involved in using the cumulative total method are as follows:

1. Calculate cumulative totals of sizes of the units ($T_i = T_{i-1} + X_i$, i=1,2,...N);
2. Choose a number R at random from 1 to T_N (=X);
3. Select U_i if $T_{i-1} < R \leq T_i$; The probability $P(U_i)$ of selecting the i-th unit is given by $P(U_i) = (T_i - T_{i-1})/ T_N = X_i /X$, which shows that the required probability of selection is achieved.
4. For selecting a sample of size n with PPS with replacement, the above operation is to be repeated n times.

The main disadvantage of this method is that it involves calculation of cumulative totals of the sizes and writing down the cumulative totals which is time consuming and costly when N is large.

Example 5.1 There are 10 villages from which a sample of size 3 is to be taken with PPS, with replacement, the measure of size being the village population. The population totals are given in column 2 of the table below. Also given below are some random numbers that can be used in drawing the sample.

2952 6641 3992 9792 7969 5911 3170 5624
4167 9524 1545 1396 7203 5356 1300 2693

Table for drawing PPS:

Village	Size (X_i)	Cumulative Total (T_i)	Numbers associated	Selected Number
1	165	165	1-165	
2	690	855	166-855	
3	1131	1986	856-1986	
4	907	2893	1987-2893	
5	582	3475	2894-3475	2952
6	2057	5532	3476-5532	3992
7	973	6505	5533-6505	
8	692	7197	6506-7197	6641
9	1738	8935	7198-8935	
10	988	9923 (=X)	8936-9923	

Starting from the first row and first column of the given random numbers, we choose three random numbers between 0001 and 9923. The first number chosen is 2952 and hence the first unit selected is U_5 as 2952 lies between 2894 and 3475; similarly, the other two units chosen are U_8 (corresponding to the random number 6641) and U_6 (corresponding to the random number 3992). Thus the PPS sample consists of the units (U_5, U_6, U_8).

(ii) **Lahiri's Method:** Lahiri suggested a method of PPS selection, which does not require calculation of cumulative totals of the sizes at all. The steps in selection of PPS sample by Lahiri's Method are as follows:

1. Select a number at random from 1 to N (say i);
2. Select another number at random from 1 to M (say, R), where M is the maximum of the sizes of the N units or some convenient number greater than the maximum size;
3. Select U_i if $R \leq X_i$; and
4. Reject U_i and repeat the above process if $R > X_i$. For selecting a sample of n units with PPS with replacement, the above procedure is repeated till n units are selected.

In the above example, N=10 and M=2057, the maximum of the sizes of 10 units. One number is now drawn at random between 1 and 10 say, 3. Now let the random number selected between 1 and 2057 be 0358 (R). Now since $R=358 \leq X_3$, we select U_3. Now suppose the second random number drawn between 1 and 10 is 2 and the number (R) selected between 1 and 2057 is 2000. Now since $R > X_2$ (=690), we reject R. and repeat the whole operation again until we get a sample of size 3.

The main advantage of this method is that it does not require writing down the cumulative totals for each unit.

Estimate of the population total: If in a population there are N units and if y_i and x_i be, respectively, the variate value and a measure of size of the i-th sampled individual with p_i being the initial probability of the unit selected at the i-th draw (i=1,2,....n), then an unbiased estimator of the population total Y is given by $\hat{Y} = (\sum y_i / p_i)/n$.

Stratified Sampling

Stratified sampling consists in classifying the population units into a certain number of groups, called strata, and then selecting samples independently from each group or stratum. An appropriate estimator for the population as a whole is obtained by suitably combining the stratum-wise estimators of the characteristic under consideration. Auxiliary information (past data or some other information) related to the character under study may be used to divide the population into various groups such that (i) units within each group are as homogeneous as possible and (ii) the group means are as widely different as possible. Thus a population consisting of N units is divided into k relatively homogeneous mutually disjoint (non-overlapping) sub-groups, termed as strata, of sizes $N_1, N_2,...N_k$ such that $N=\sum N_s$. If a random sample of size n_s (s=1,2,....k) is drawn from each of the stratum respectively such that $n=\sum n_s$, the sample is termed as stratified random sample of size n and the technique of drawing such a sample is called stratified random sampling.

The criterion which enables us to classify various sampling units into different strata is termed as stratifying factor. Some of the commonly used stratifying factors are age, sex, educational or income level, geographical area, economic status and so on. A stratifying factor is said be effective if it divides the given population into different strata which are homogeneous within themselves and units in different strata are as heterogeneous as possible. Some examples of stratification are given below:

(i) Division of the whole area of a region into coastal, plains and hilly areas;
(ii) Classification of retail stores in a city on the basis of products sold;
(iii) Classification of non-agricultural enterprises as per the activity pursued;
(iv) Grouping of rural households by size class of land possessed;
(v) Grouping of urban households by monthly per capita expenditure.

Advantages of stratified sampling

(i) The main advantage of using stratified sampling is the possible increase in efficiency per unit of cost in estimating the population characteristics.
(ii) There is considerable flexibility in stratified sampling in the sense that the sampling and the estimation procedures may differ from stratum to stratum depending on the nature of supplementary information available and that the demarcation of the strata boundaries and the allocation of the total sample size to the strata may be done so as to make the estimator most efficient from the point of view of sampling variability and cost.
(iii) Use of stratification is of considerable importance in case of skewed population, since greater weightage may have to be given for the few extremely large units for reducing the sampling variability.
(iv) In an unstratified sample some parts/groups of the population may be over-represented, while some others may be under-represented and some part may be excluded altogether. Stratified sampling ensures any desired representation in the sample of the various strata in the population. Possibility of completely excluding any essential group of the population is ruled out in stratified sampling. It thus gives a more representative cross section of the population and is frequently regarded as the most efficient system of sampling.

(v) If the sampling frame is available in the form of sub-frames, which may be for regions or from specified categories of units, it may be operationally convenient and economical to treat the sub-frames as strata for sample selection.

(vi) The amount of supplementary information available, and hence the methods of selection and estimates to be used, may differ from region to region and also from one group of population to another. In such cases using these regions or groups as strata is not only operationally convenient, but also saves considerable amount of money and time. Thus Stratified sampling has more administrative convenience.

Allocation of sample size:

Intuitively it may be felt that the allocation of the sample size to the strata would depend on the stratum sizes and the within-strata variation. If the cost per unit is presumed to be constant for all the strata, the stratum or strata accounting for a substantial part of the variation should receive a larger allocation. On the other hand, if the contribution from each stratum to the sampling variance is almost constant, then the stratum or strata where the cost of survey per unit is large should get a smaller allocation.

Optimum allocation:

The problem of optimum allocation consists in determining the number of units to be selected from the different strata with a view to minimizing (i) the sampling variance for a given cost, or (ii) the cost of the survey while ensuring a specified value for the sampling variance. In both the cases it may be shown that

$n_s \propto N_s * \sqrt{(V_s/C_s)}$ where n_s = number of units to be selected from s-th stratum (i.e sample size of the s-th stratum), N_s = Number of units (in the population) in the s-th stratum (i.e. population size in the s-th stratum), V_s = sampling variance of the s-th stratum and C_s = average cost of surveying one unit in the s-th stratum.

In particular, if $C_1 = C_2 = \ldots = C_k$, finding the cost is equivalent to fixing the total sample size i.e. making $n_1+n_2+\ldots+n_k=n$. In that case, the optimum values of n_s ($s=1,2\ldots k$) are given by $n_s = n^*[(N_s * S_s)/\sum (N_s * S_s)]$, where $S_s = \sqrt{V_s}$, sampling standard deviation of the s-th stratum. This is called Neyman's formula for optimum allocation.

The main difficulty in having the optimum allocation is that it requires knowledge of V_s which information is usually either not available or difficult to obtain. But approximation to this allocation may be obtained by using estimates of V_s from a previous survey or by conducting a preliminary survey or pilot survey, which is a survey of a smaller scale.

Proportional allocation:

When no other information except $\{ N_s \}$ is available, the allocation of a given sample size n may be done in proportion to N_s provided there is evidence to expect that the sampling variance in the smaller strata is less than that in the larger ones.

In this allocation, $n_s = n^* N_s /N$. This allocation was proposed by Bowley and is popularly known as 'Bowley's formula for proportional allocation'. This procedure of allocation is often resorted to in practice because of its simplicity.

Gain due to stratification:

It may be proved mathematically that Stratified SRS is definitely more efficient than unstratified SRS whenever n_s is taken proportional to (i) N_s or (ii) $N_s * \sigma_s^2$ where σ_s^2 =

population variance of s-th stratum. It may be mentioned here that stratification is not justified if (i) stratification itself is too costly and/or (ii) the between strata variance is not large enough to effect a sufficient gain in the accuracy as compared to SRS.

Estimate of population mean: suppose a population of N units is divided into k strata. Let N_s be the number of units in the s-th stratum, and let Y_{si} be the value of the study variable for the i-th unit in the s-th stratum.
Then the population mean $\bar{Y}$ can be expressed as $\bar{Y}=\sum W_s * \bar{Y}_s$, where $W_s = N_s/N$ and $\bar{Y}_s = \sum Y_{si} / N_s$, the population mean for the s-th stratum.

An unbiased estimator of $\bar{Y}$ can be obtained by estimating unbiasedly the stratum means $\{ \bar{Y}_s \}$ on the basis of random samples drawn from each stratum with sampling schemes, which need not necessarily be the same for all strata. If the sampling scheme is SRS for all the strata, then $\bar{y}_s$, the sample mean for the s-th stratum is an unbiased estimator of $\bar{Y}_s$ and hence an unbiased estimator for population mean $\bar{Y}$ is given by $\sum W_s * \bar{y}_s$, where $W_s = N_s/N$.

Example: Suppose in a socio-economic survey 2000 households in a particular region of a city are stratified according to their monthly income. The number of households (N_s) in each stratum are given in the following table. Using Bowley's formula for proportional allocation, determine the size (n_s) of the samples to be drawn from each stratum if the total sample size (n) is 100.

Stratum no.	Monthly Income (Rs.)	No. of households
1	Below 2000	400
2	2001-5000	500
3	5001-10,000	700
4	10,001 and above	400

Using Bowley's formula for proportional allocation, the sample size for the s-th sample is given by $n_s = n * N_s/N$, where N_s is the s-th stratum size, N is population size and n= sample size. In this case, N=2000, n=100. Hence n_1=100*400/2000=20; n_2=100*500/2000=25; n_3=100*700/2000=35 and n_4=100*400/2000=20.

Systematic Sampling

Systematic sampling is a commonly employed technique if the complete and up-to-date list of the sampling units is available. Systematic sampling is operationally more convenient than simple random sampling, and at the same time ensures for each unit equal probability of inclusion in the sample.

The technique of systematic sampling consists in selecting every k-th unit starting with the unit corresponding to a number r chosen at random from 1 to k, where k is taken as the integer nearest to N/n, the reciprocal of the sampling fraction where N is the population size and n is the sample size. The random number r chosen from 1 to k is known as the random start and the constant k is termed as the sampling interval. A sample selected by this procedure is termed a systematic sample with a random start.

The selection of every k-th strip in forest sampling for the estimation of timber, the selection of corn fields every k-th mile apart for observation on incidence of borders, selection of every k-th household or k-th village from a list of households or villages, after the first unit is chosen with the help of random numbers less than or equal to k, are all examples of systematic sampling schemes.

It may be seen that in systematic sampling, the value of r determines the whole sample. In other words, this procedure amounts to selecting with equal probability one of the k possible groups of units (samples) into which the population can be divided in a systematic manner. Hence this is a case of mixed sampling which is partly probabilistic and partly non-probabilistic. This is probabilistic because the first member of the sample is drawn at random (with equal probabilities) from the first k units and non-probabilistic since the other members in the sample are fixed by the choice of the first member. In drawing a systematic sample of size n from a population of size N one may encounter any of the two cases (i) when $N/n=k$ is an integer i.e. $N=n*k$ and (ii) when $N\neq n*k$. Let us discuss both the cases one by one. Notationally, the units in a systematic sample of n units drawn from a population of N units are given by

$$\{U_{r+jk}\},\ j=0,1,2,\ldots,\ r+(n-1)k$$

Case (i) $N=n*k$ i.e. $K=N/n$ is an integer. (Linear Systematic sampling)

If N is a multiple of n, then the number of units in each of the k possible systematic samples is n. In this case systematic sampling amounts to grouping the N units into k samples of exactly n units each in a systematic manner and selecting one of them with probability 1/k. Systematic sample drawn in this manner is called a linear systematic sample. The k possible samples corresponding to k possible random start are given in the Table 6.1. From this table it is clear that each of the N units occurs once and only in one of the k samples, thus ensuring equal probability of inclusion in the sample for every unit in the population.

Table 6.1. Systematic sample of n units from N (=n*k) units

Random start	Sample composition	Probability
1	$U_1\ U_{1+k}\ldots U_{1+jk}\ldots U_{1+(n-1)k}$	1/k
2	$U_2\ U_{2+k}\ldots U_{2+jk}\ldots U_{2+(n-1)k}$	1/k
.	.	.
.	.	.
r	$U_r\ U_{r+k}\ldots U_{r+jk}\ldots U_{r+(n-1)k}$	1/k
.	.	.
.	.	.
k	$U_k\ U_{2k}\ldots U_{(j+1)k}\ldots U_{nk}$	1/k

In order to select a linear systematic sample, one needs to follow these steps:

1. Number the units in the frame from 1 to N (where N is the total population size).
2. Determine the sampling interval (k) by dividing the number of units in the population by the desired sample size. For example, to select a sample of size 4 from a population of size 20, the sampling interval is $20 \div 4 = 5$ i.e. $k = 5$.
3. Select a random number between 1 and k. This number would be the first number included in the sample. Using a random number table, select a number between 1 and 5. If it is 3, then third unit in the frame would be the first unit included in the sample; if the first number chosen is 4, the sample would start with the fourth unit of the frame.
4. Select every k-th (in this case every fifth) unit after that first number. For example, if the random start is 3, the sample will consist of the following units to make up a sample of four: 3, 8, 13, and 18 (i.e. U_3, U_8, U_{13}, and U_{18}) and if the random start is 4, it would consist of the units 4, 9, 14 and 19 (i.e. U_4, U_9, U_{14}, and U_{19}).

In case of linear systematic sampling, the sample mean provides an unbiased estimate of the population mean.

Case (ii) $N \neq n*k$ i.e. $K=N/n$ is not an integer (Circular systematic sampling)

If N is not a multiple of n, then the number of units selected systematically with the sampling interval k equal to the integer nearest to N/n need not necessarily be equal to n, the required sample size. This problem may be overcome by adopting a device, known as circular systematic sampling. This device consists in choosing a random start from 1 to N and selecting the unit corresponding to this random start and thereafter every k-th unit in a cyclical manner till a sample of size n is obtained, k being the integer nearest to N/n.

That is, if r is a number selected at random from 1 to N, the sample consists of the units corresponding to the numbers

$$\text{and} \quad \left.\begin{array}{ll} \{r+j*k\} & \text{if } (r+j*k) \le N \\ \{r+j*k-N\} & \text{if } (r+j*k) > N ; \end{array}\right\} \quad j=0,1,2,\ldots..(n-1). \quad \ldots\ldots\ldots \ (**)$$

This technique ensures equal probability of inclusion in the sample for every unit.

For example, to select a sample of size 3 from a population of size 11, we may adopt this technique. Here k= integer nearest to $N/n = 4$. Using a random number table, select a number between 1 and N=11. This number would be the first number included in the sample. If it is 3, then third unit of the frame would be the first unit included in the sample. Thereafter choose every 4^{th} unit in a cyclical manner till a sample of size 3 is obtained. Thus the sample will consist of the units having number 3, 7 and 11 i.e. the unitsU_3, U_7, U_{11}. If the random start is 5, then the fifth unit of the frame would be the first unit included in the sample and the other units will be unit numbers 9 (=5+4) and 2 (= 5+2*4-11) (following ** above) i.e. the units U_5, U_9, and U_2.

As in the case of linear systematic sampling, in circular systematic sampling also the sample mean is an unbiased estimator of the population mean.

The method of systematic sampling is extensively used on account of its low cost and simplicity in the selection of the sample. Since the actual procedure of systematic sampling is quite simple, it would be easy to train persons in using and hence it may be desirable to adopt this procedure whenever the sampling work has to be carried out by a large number of persons stationed in different areas. A systematic sample also offers great advantages in organizing control over field work.

In systematic sampling, the relative position in the population of the different units included in the sample is fixed. There is consequently no risk in the method that any

large contiguous part of the population will fail to be represented. Indeed, the method will give an evenly spaced sample and is, therefore, likely to give a more precise estimate of the population mean than random sampling would, unless the k-th units constituting the sample happen to be alike or correlated.

The main difference between the procedures of systematic sampling and simple random sampling is that enumeration of all possible samples and selection of one of them with equal probability are much simpler in the former than in the latter procedure. It is much easier and quicker to draw a systematic sample and the work may be done by layman. Apart from this operational convenience, which is of considerable importance in large-scale sampling work, this sampling procedure provides estimators more efficient than those provided by SRS under certain conditions usually met with in practice.

Systematic sampling closely resembles cluster sampling, a systematic sample being equivalent to a sample of one cluster selected out of k clusters of n units each.

Systematic sampling has particularly been found useful in forest surveys, where survey of randomly selected area units would be time consuming and costly due to difficulties of access and identification. Systematic sampling is also widely used in population or other censuses for collecting detailed information or for quick tabulation of results.

A major drawback of systematic sampling is that it does not provide an unbiased estimate of the sampling error. Another disadvantage of systematic sampling is that it may yield highly biased estimates if there are periodic features associated with the sampling interval, i.e. if the frame has a periodic feature and k is equal to or a multiple of the period.

Cluster/Multistage Sampling

Any sampling procedure presupposes the division of the population into a finite number of distinct and identifiable units called the sampling units. The smallest units into which the population can be divided are called the elements of the population, and groups of elements the clusters. When the sampling unit is a cluster, the procedure of sampling is called cluster sampling.

Cluster sampling consists in forming suitable clusters of units and surveying all the units in a sample of clusters selected according to an appropriate sampling scheme. Thus in this case the total population is divided, depending on problem under study, into some recognizable subdivisions which are termed as clusters. The advantages of cluster sampling from the point of view of cost arise mainly due to the fact that collection of data for nearby units is easier, faster, cheaper and more convenient than observing units scattered over a region. For instance, in a population survey it may be cheaper to collect data from all persons in a sample of households than from a sample of the same number of persons selected directly from all the persons. Similarly, it would be operationally more convenient to survey all households situated in a sample of areas such as villages than to survey a sample of the same number of households selected at random from a list of all households. Another example of the utility of cluster sampling is provided by crop surveys, where locating a randomly selected farm or plot requires a considerable part of the total time taken for the survey, but once the plot is located, the time taken for identifying and surveying a few neighbouring plots will generally be only marginal.

Because of its operational convenience and the possible reduction in cost, cluster sampling is resorted to in many surveys, using mutually exclusive clusters formed by grouping the nearby units or units which can be conveniently observed together. In general, for a given total number of sampling units, cluster sampling is less efficient than sampling of individual units from the view point of sampling variance, as the later is expected to provide a better cross section of the population than the former due to the usual tendency of units in a cluster to be similar. In fact, the sampling efficiency of cluster sampling is likely to decrease with increase in cluster size. However, cluster sampling is operationally more convenient and less costly than sampling of units directly due to the possible saving in time for journey, identification, contact, etc. and hence in many practical situations the loss in sampling efficiency is likely to be offset by the reduction in cost.

In a general sense, any system of sampling may be regarded as a kind of cluster sampling, since in every sampling scheme the units are conceptually grouped to form samples (clusters) and one of them is selected with a certain probability. For example, systematic sampling may be considered a particular case of cluster sampling, since in this case the population is divided into a number of clusters, each cluster consisting of units distributed at a fixed interval over the whole population and one such cluster is selected at random. But, by cluster sampling is usually meant sampling of clusters of units formed by grouping neighbouring units or units which can be conveniently surveyed together. It may be mentioned here that the various sampling schemes like SRS, systematic sampling, PPS and stratified sampling can be applied to sampling of clusters by treating the clusters themselves as sampling units.

Optimum Cluster size

It may be seen that for a given total sample size in terms of units, the sampling variance increases with cluster size and decreases with increasing number of clusters, while the cost decreases with increasing cluster size and increases with number of clusters. Hence,

in practice, it is necessary to strike a balance between these two opposing points of view by finding the optimum values for the cluster size and the number of sample clusters, which would minimize the sampling variance for a fixed cost or alternatively minimize the cost for a specified sampling variance.

Estimator of population mean:

Suppose a finite population of N*M units is divided into N mutually exclusive clusters of M units each and one cluster is selected with SRS for estimating the population mean. Then the sample cluster mean gives an unbiased estimate of the population mean.

Multi-stage sampling:

Though cluster sampling is more economical under certain circumstances, it is generally less efficient than sampling of individual units directly. A compromise between cluster sampling and direct sampling of units can be achieved by selecting a sample of clusters and surveying only a sample of units in each sample cluster instead of completely enumerating all the units in the sample clusters. Such a procedure is known as two-stage sampling, since the units are selected in two stages. This procedure may be generalized to give rise to multi-stage sampling.

In multi-stage sampling, the material to be sampled is regarded as composed of a number of first-stage (primary) sampling units (fsu), each of which is made up of a number of second stage (or secondary) units (ssu), each of which, in turn, is made up of a number of third stage (or tertiary) units and so on, until we reach the ultimate sampling units in which we are interested. The sampling is also carried out in stages. At the first stage, the first stage sampling units are sampled by some suitable random method. At the second stage, a sample of second stage units is selected from each of the selected first stage-units, again by some suitable random method. Further stages may be added, if necessary, to get a sample of the ultimate sampling units. For example, to get a sample of crop-fields growing paddy in West Bengal, one may first get a sample of districts, then a sample of villages from each of the selected districts and finally a sample of crop-fields from each selected village.

This procedure, being a compromise between uni-stage or direct sampling of units and cluster sampling can be expected to be (i) more efficient than direct sampling and less efficient than cluster sampling from the view-point of operational convenience and cost and (ii) less efficient than direct sampling and more efficient than cluster sampling from considerations of sampling variability, when the sample size in terms of number of ultimate units is fixed. Multi-stage sampling introduces more flexibility into the sampling procedure. It enables existing divisions and sub-divisions of the material to be taken as sampling units at different stages. The construction of a second-stage frame is necessary only for the selected first stage units. This means a great savings in the operational costs, particularly if the survey covers a large area including under-developed pockets.

It may be mentioned that multi-stage sampling may be the only feasible procedure in a number of practical situations, where a satisfactory sampling frame of ultimate observational units is not readily available and the cost of obtaining such a frame is prohibitive or where the cost of locating and physically identifying the ultimate stage units (usu) is considerable. For example, for conducting a socio-economic survey in a region, where generally household is taken as the usu, a complete and up-to-date list of all the households in the region may not be available, whereas a list of villages or urban blocks which are groups of households in rural and urban areas respectively is available. In such a case, a sample of households may be drawn from each of selected village and urban block after making a complete list of households in them.

In practice, it usually happens that we have more information for groups of sampling units than for individual units. Hence, if these groups are taken as fsu's, the information available for them can be used in effecting good stratification and in selecting the sample of fsu's. Further, since the ssu's are selected only from the sample of fsu's, it would be practicable to collect some suitable information about the ssu's at the time of listing them and use this information for obtaining a better sample of ssu's. Because of this, it may be possible that a multi-stage design, where the information available at every stage is properly utilized, is more efficient than the uni-stage sampling from the point of view of sampling variability.

Estimator of population total:

Let us consider the simple case of SRS at both the stages of a two-stage sampling scheme. Suppose a sample of n fsus is selected with SRS and from the i-th selected fsu a sample of m_i ssus is selected again with SRS. An unbiased estimator of population total Y is given by

$$\hat{Y} = (N/n) * \sum (M_i/m_i) * \sum y_{ij}$$

Where y_{ij}= value of the j-th selected ssu in the i-th selected fsu ($j=1,2\ldots m_i$; $i=1,2,\ldots n$)

CONCEPTS, DEFINITIONS AND CLASSIFICATIONS

A number of accounting, economic and other terms related to the factories are used in conduct of Annual Survey of Industries. For proper compilation/ verification of ASI return, it is essential that the staff engaged in the conduct of ASI properly understood the terminology used in ASI. It is in this context that the concepts and definitions of some of the important terms are given below.

ACCOUNTING YEAR

For the purpose of ASI, the accounting year is the period on which the factory closes its books of account. With the enactment of Income Tax Act, by and large, the accounting year of all factories is from April to March.

BASIC MATERIALS

Basic materials are the materials which are important and of key nature to the industry on which the manufacturing process is based, vig. metal for machine, leather for shoe. Such material is not lost through the process of production but only changes its forms.

BONUS

Profit sharing bonus, festival bonus, year-end bonus, and all other bonuses and ex-gratia payments paid at less frequent intervals are covered by this term.

COMPENSATION OF EMPLOYEES

Compensation of employees is the total of emoluments and supplement to emoluments.

CONSUMABLE STORES

All such materials which assist the manufacturing process and loose their identity without entering the products are called consumable stores, e.g., cotton waste.

CONTRACT WORKER

All persons who are not employed directly by an employer but through the third agency, i.e. contractor, are termed as contract workers. Those workers may be employed with or without the knowledge of the principal employer.

DEPRECIATION

Depreciation is consumption of fixed capital by the factory due to wear and tear and obsolescence during the accounting year and is taken as provided by the factory owner, or if not provided by the factory this is estimated on the basis of cost of installation and working life of the fixed assets.

EMPLOYEES

Employees relate to all persons engaged by the factory whether for wages or not, in work connected directly or indirectly with the manufacturing process and include all administrative, technical and clerical staff as also labour in production of capital assets for

factory's own use. This is inclusive of persons holding position of supervision or management or engaged in administrative office, store-keeping section and welfare section, watch and ward staff, sales department as also those engaged in the purchase of raw materials etc. and production of fixed assets for the factory. It also includes all working proprietors and their family members who are actively engaged in the work of the factory even without any pay and the unpaid members of the co-operative societies who work in or for the factory in any direct and productive capacity. Persons in the head office connected with the manufacturing activity of the factory are also included in this item.

EMOLUMENTS

These are defined in the same way as wages but paid to all employees plus imputed value of benefits in kind i.e. the net cost to the employers on those goods and services provided to employees free of charge or at markedly reduced cost which are clearly and primarily of benefit to the employees as consumers. It includes profit sharing, festival and other bonuses and ex-gratia payments paid at less frequent intervals (i.e. other than bonus paid more or less regularly for each period). Benefits in kind include supplies or services rendered such as housing, medical, education and recreation facilities. Personal insurance, income tax, house rent allowance, conveyance etc. for payment by the factory also is included in the emoluments.

FACTORY

Factory is one, which is registered under Sections 2m(i) and 2m(ii) of the Factories Act, 1948. The Sections 2m(i) and 2m(ii) refer to any premises including the precinct thereof (i) wherein ten or more workers are working, or were working on any day of the preceding twelve months, and in any part of which a manufacturing process is being carried on with the aid of power, or is ordinarily so carried on, or (ii) wherein twenty or more workers are working, or were working on any day of the preceding twelve months and in any part of which a manufacturing process is being carried on without the aid of power, or is ordinarily so carried on. Closed factories with fixed assets on site are also considered as registered factories till they are de-registered and removed from the live-register maintained by the Chief Inspector of Factories (CIF) in the State.

FINISHED GOODS

Finished Goods are those, which are manufactured by the factory for sale. Finished goods should conform to a prescribed standard.

FIXED CAPITAL

Fixed Capital represents the depreciated value of fixed assets owned by the factory as on the closing day of the accounting year. Fixed assets are those, which have normal productive life of more than one year. Fixed capital covers all type of assets, new or used or own constructed, deployed for productions, transportation, living or recreational facilities, hospitals, schools, etc. for factory personnel. It would include land, building, plant and machinery, transport equipment etc. It includes the fixed assets of the head office allocable to the factory and also the full value of assets taken on hire-purchase basis (Whether fully paid or not) excluding interest element. It excludes intangible assets and assets solely used for post-manufacturing activities such as, sale, storage, distribution, etc.

FUEL CONSUMED

Fuel Consumed represent total purchase value of all items of fuels, lubricants, electricity, water (purchased to make steam) etc. consumed by the factory during the accounting year except those which directly enter into products as materials consumed. It excludes that part of fuels, which is produced and consumed by the factory in manufacture i.e., all intermediate products and also fuels consumed by employees as part of amenities. It includes quantities acquired and consumed from allied concerns, their book value being taken as their purchase value and also the quantities consumed in production of machinery or other capital items for factory's own use.

GROSS OUTPUT

Gross output is defined to include the ex-factory value, (i.e., exclusive of taxes, duties, etc. on sale and inclusive of subsidies etc., if any) of products and by-products manufactured during the accounting year, and the net value of the semi-finished goods, work-in-process, (represents the excess/deficit of value of semi-finished goods or work-in-process at the end of the accounting year over that of the beginning of the year plus net balance of semi-finished fixed assets on factory's capital account) and also the receipts for industrial and non-industrial services rendered to others, value of semi-finished goods of last year sold in the current year and sale value of goods sold in the same condition as purchased. Value of gross output and total output has been used in the text inter-changeable to mean the same thing.

GROSS VALUE OF PLANT AND MACHINERY

Gross value of plant and machinery represents the total original (undepreciated value of installed plant and machinery as at the end of the accounting year. It includes the book value of own constructed plant and machinery, if installed, and the approximate value of rented-in plant and machinery as at the time of renting in but excludes the value of rented-out plant and machinery. Total value of all the plant and machinery acquired on hire-purchase basis is also included. Thus it represents the gross value of plant and machinery engaged in production process.

INDUSTRIAL SERVICES

Any services taken or rendered from one to another unit resulting in increase in the value of material during the manufacturing process are industrial services.

INTERMEDIATE PRODUCT

Intermediate Product is a product which is obtained during a manufacturing process, which may or may not be saleable and is not the intended final product.

NVESTED CAPITAL

23. Invested capital is the total of fixed capital and physical working capital.

LABOUR TURNOVER

Labour turnover measures the extent of change in the working force due to accession and separation during a given period. The term 'accession' was defined as the total number of workers added to employment during the period, whether new or re-employed or transferred from other establishments or units under the same management. Inter-departmental transfers within the same establishment are, however ignored. The term 'separation' implies termination of employment at the instance of worker or employers. It includes termination of services due to death or retirement. As in the case of accession, transfers to other establishments are included but transfers within the same establishment are ignored. Retrenchment as a result of rationalisation or modernisation or any other cause, is also treated as separation.

MANDAYS WORKED

These are obtained by summing up the number of mandays worked by persons working in each shift over all the shifts on all days, i.e. both manufacturing and non-manufacturing days. This figure excludes persons who are paid but remain on leave, strike, etc.

MANDAYS PAID FOR

The number of mandays paid for is arrived at by summing up the number of employees paid for in each shift. This also includes mandays on weekly schedule holidays if paid for and those absences with pay as also mandays lost through lay off/ strike for which compensation was payable.

MANUFACTURING PROCESS

As per Section 2(k) of the Factories Act, 1948, manufacturing process means any process for :

i) Making altering, repairing, ornamenting, finishing, packing, oiling, washing, cleaning, breaking up, demolishing, or otherwise treating or adopting any article or substance with a view to its use, sale, transport, delivery or disposal, or
ii) Pumping oil, water, sewage, or
iii) Generating, transforming or transmitting power, or
iv) Composing types for printing, printing by letter press, lithography, photogravure or other similar process, or book-binding, or
v) Constructing, re-constructing, repairing, finishing, breaking up ships or vessels or
vi) Preserving or storing any article in Cold Storage.

MATERIALS CONSUMED

Materials consumed represent the total delivered value of all items of raw materials, components, chemicals, packing materials and stores which actually entered into the production process of the factory during the accounting year. It also includes the cost of all the materials used in the production of fixed assets, including construction work for factory's own use. Components and accessories fitted as purchased with the finished product during the accounting year are also to be included. It excludes intermediate products.

Intermediate products in the above context mean all those products which are produced by the factory and consumed for further manufacturing process.

NET VALUE ADDED

This is the increment to the value of goods and services that is contributed by the factory and is obtained by deducting the value of total inputs and depreciation from gross value of output.

NET VALUE OF SEMI-FINISHED GOODS

It represents the excess/deficit of value of semi-finished goods and/or goods-in-process at the end of the accounting year over that at the beginning of year.

NON-INDUSTRIAL SERVICES

All such services which do not have a direct bearing on the manufacturing process but are needed by any manufacturing unit are called non-industrial services, say, transport.

NON WORKING DAY

Apart from manufacturing day and repair and maintenance days there may be some non-working days. Non working days are those days on which the workers give their attendance but due to non-availability of raw materials power etc. no effective work is done. As the workers are paid for these days such days are also taken into account for the purpose of labour statistics.

OUTSTANDING LOANS

Outstanding loans represent all loans, whether short-term or long-term, whether interest bearing or not, outstanding according to the books of the factory as on the closing day of accounting year.

PHYSICAL WORKING CAPITAL

This is defined to include all physical inventories owned, held or controlled by the factory as on the closing day of the accounting year such as the materials, fuels and lubricants, stores, etc. that enter into products manufactured by the factory itself or supplied by the factory to others for processing. Physical working capital also includes the value of stock of materials, fuels and stores etc. purchased expressly for re-sale, semi-finished goods and goods-in-process on account of others and goods made by the factory which are ready for sale at the end of the accounting year. However, it does not include the stock of the materials, fuels, stores, etc. supplied by others to the factory for processing. Finished goods processed by others from raw materials supplied by the factory and held by them are included and finished goods processed by the factory from raw materials supplied by others, are excluded.

PRODUCTS

These are defined to include the ex-factory value (i.i. exclusive of taxes, duties etc. on sale and inclusive of subsidies etc., if any) of all products and by-products, excluding intermediate products, that have been completed during the accounting year for sale whether actually sold during the accounting year or entered into books. Also include fixed assets produced by the factory for its own use.

PRODUCTIVE CAPITAL

This is the total of fixed capital and working capital.

REFERENCE PERIOD

For ASI it is corresponding financial year. For example, for ASI 2003-2004 the reference period is the financial year commencing from 1st April 2003 and ending on 31st March 2004 or the accounting year of the factory ending on any date between 1.4.03 to 31.3.04.

SUPPLEMENT TO EMOLUMENTS

These include:

i) Employer's contribution to old age benefits, i.e., provident fund, pension, gratuity, etc.
ii) Employer's contribution towards other social security charges such as Employees' State Insurance, compensation for work injuries, occupational diseases, maternity benefits, retrenchment and lay-off benefits etc. and
iii) Group benefits like direct expenditure on maternity, creches, canteen facilities, educational, cultural and recreational facilities and grant to trade unions, co-operative stores etc. meant for employees.

SURVEY PERIOD

Survey period is a period during which work of any Annual Survey of Industries is undertaken. Since ASI 1978-79 the survey period has been fixed from 1st July to 30th June of the next year. From ASI 1998-99 the survey period has been changed from 12 months to 4 months i.e. from Nov'99 to Feb'2000. The survey period for ASI 99-00 has been again changed from October to March. The survey period for ASI 2000-2001 is from Sept. 2001 to April, 2002.

TOTAL INPUT

This comprises gross value of fuel materials etc. consumed (as defined above) and also other inputs viz. (a) cost of non-industrial services received from others (b) cost of materials consumed for repair and maintenance of factory's fixed assets including cost of work done by others to the factory's fixed assets (c) cost of contract and commission work done by others on materials supplied by the factory (d) cost of office supplies and products reported for sale during last year & used for further manufacture during the accounting year.

WAGES

Wages are defined to include all remuneration capable of being expressed in monetary terms and also payable paid more or less regularly in each pay period to workers (defined above) as compensation for work done during the accounting year. It includes;

a) Direct wages and salary (i.e. basic wages/salaries, payment of overtime, dearness, compensatory, house rent and other allowances.
b) Remuneration for period not worked (i.e. basic wages), salaries and allowances payable for leave period, paid holidays, lay-off payments and compensation for unemployment (if not paid from source other than employers).
c) Bonus and ex-gratia payment paid both at regular and less frequent intervals (i.e. incentive bonuses and good attendance bonuses, production bonuses, profit sharing bonuses, festival or year end bonuses etc.). It excludes layoff payments and compensation for employment except where such payments are this purpose i.e. payments not made by the employer. It excludes employer's contribution to old age benefits and other social security charges, direct expenditure on maternity benefits and creches and other group benefit in kind and travelling and other expenditure incurred for business purposes and reimbursed by the employer. The wages are expressed in terms of gross value i.e. before deductions for fines, damages, taxes, provident fund, employee's state insurance contribution etc. Benefits in kind (perquisites) of individual nature are only included.

WORKING CAPITAL

Working Capital is the sum total of the physical working capital as already defined above and the cash deposits in hand and at bank, land, the net balance of amounts receivable over amounts payable at the end of the accounting year. Amounts receivable include value of credit items on revenue account, such as sums due to the factory for goods sold, amounts advanced in connection with normal factory work, bills of exchange payable to the factory, payments made in advance such as for fire insurance, telephone charges, rates and taxes, call deposits and security deposits having a normal life of less than one year, etc. It excludes unused overdraft facility, fixed deposits irrespective of duration, advances for acquisition of fixed assets, long-term loans including interest thereon and investment.

WORKING DAY

Working day means the days on which a manufacturing process and/ or repair or maintenance work was carried on.

WORKMEN AND STAFF WELFARE EXPENSES

These include expenditure incurred by the employer on the maternity benefits and crèches and other benefits such as supply of food, beverages, tobacco, clothing and group lodging at confessional rates and educational, cultural and recreational facilities and services and grants to trade unions and cooperative stores meant for employees. All group benefits are included.

INTRODUCTION

Health Statistics deals with application of statistical methods to varied information of public health importance.

Medical Statistics deals with application of statistical methods to the study of disease, disability, efficacy of vaccine, a new regime etc.

Vital Statistics is the ongoing collection by government agencies of data relating to vital events such as births, deaths, marriages, divorces and health and disease related states and events, which are deemed reportable by local health authorities.

Biostatistics can be defined as an art and science of collection, compilation, presentation, analysis and logical interpretation of biological data affected by multiplicity of factors. This appears to be a comprehensive definition encompassing health, medical and vital statistics together with statistical methods encountered.

Biostatistics helps to define what is normal and it also gives limits of normality. In statistical terms one must realize that 'normal' means 'usual' and not its literal meaning. Take example of haemoglobin (Hb) level. After estimating Hb level in many individuals one can come to a conclusion that normal range of haemoglobin in adult males is 13 to 15 gm%. Thus it tells us the physiological limits and pathological outcome like anaemia when the deviation is beyond normal limits.

Bio statistical methods provide us demographic profile of any country, indicators of mortality, morbidity and fertility in the population, information regarding rise and fall of diseases and development in health sector etc.

The principal applications of statistics in public health are

1) Population estimation and forecasting
2) Surveys of population characteristics, health needs and problems
3) Analysis of health trends
4) Epidemiologic research
5) Programme evaluation
6) Programme planning
7) Budget preparation and justification
8) Operational and administrative decision making
9) Health education

Thus it can be concluded that, biostatistics is a useful aid for medical professionals but it is also a double-edged weapon. If used indiscriminately it can spoil the show but if used judiciously it can do wonders and add flying colours to our performance as teaching, treating or testing doctors.

TOOLS FOR MEASUREMENT

There are three basic tools of measurement. These are-

- Rate
- Ratio and
- Proportion

Rate

In the earthquake at Killari 9000 people lost their lives where as in another earthquake in Peru 1,000 people died. Such information is probably adequate for a government, which is planning to provide help in kinds or cash to the kins of the deceased. However, this gives little information it we want to compare these events and associated mortality. In school, "A"

27 students developed food poisoning and in school, "B" 93 students developed food poisoning. One might be tempted to comment that more cases of food poisoning have occurred in school B than A. Now we are told that in school "A", 27 cases occurred in month of January 94 and there are 270 students in the school and in school "B", 93 cases occurred from Jan. to June of 1994, and there are 1860 students in the school. One will now realize that in school "A" in one month almost 10% (27/270) students suffered from food poisoning whereas in school "B" only 5% students suffered from food poisoning that too in 6 months (if distributed evenly over months then the figure for January 94 can be 5/6 = 0.83%). This highlights the importance of "when" i.e. period and "out of how many" whenever we want to compare. To avoid the problems mentioned above we use the tool known as rate. A rate consists of (a) numerator (how many) (b) denominator (out of how many) (c) time specification (period or time frame) and (d) multiplier. The example of a typical rate is birth rate. It is given as-

$$\text{Birth rate} = \frac{\text{Number of births in 2001}}{\text{Mid year population in 2001}} \times 1000$$

If we put hypothetical figures. (20/2000) x 1000, then the birth rate for 2001 is 1000 population. Multiplier used in above example is 1000, however as per convenience or for avoiding fractions one can use 100, 1,000 or 1,00,000 as multipliers. There are various types of rates –

Crude Rates – This rate is called crude because here denominator used is entire mid year population and we are not including specific information in numerator, e.g. in crude birth rate we include all births whether live or still births. It is difficult to use this rate for comparison with other times or places due to possibility of significant differences in age and sex compositions of various populations.

Specific Rates – These are the actual observed rates due to specific causes (e.g. AIDS, Cancer); or occurring in specific groups (e.g. under five children, women from 15 to 44 years or males); or during specific time periods (week, month or year). Mortality rate of tuberculosis, couple protection rate and 5-year survival rate are a few examples of such rates.

Standardized Rates – To make two populations and their rates comparable, adjustment either by direct or indirect method for age, sex, religion etc. can be made. Thus we can get age or sex standardized rates.

Other Rates – Where time period is not a year. Of course these rates can be converted to annual rates by using appropriate multiplier.

$$\text{a) Quinquennial rate} = \frac{\text{No. of vital events in the population during the period of five years}}{\text{population estimated of the middle of the five years}} \times 1000$$

To convert to annual rate, multiply by 1/5.

$$\text{b) Decennial rate} = \frac{\text{No. of vital events in the population during the period of one decade i.e. 10 years}}{\text{population estimated at the middle of the ten years.}} \times 1000$$

To convert to annual rate, multiply by 1/10.

c) Weekly rate = $\frac{\text{No. of vital events occurring in one week}}{\text{Total mid year estimated population}} \times 100$

To convert to annual rate, multiply by 52.

Ratio

Specialty of this tool of measurement is that it expresses relation between two unrelated quantities. e.g. in a basket there are 30 mangoes and 20 apples we can say that ratio of mangoes to apples is 3/2 or 3:2. Similarly we can get ratio or WBCs to RBCs, serum cholesterol to serum calcium etc. Here numerator is not a part of the denominator. It is expressed as a:b or a/b. Hospital bed population ratio, doctor-nurse ratio are few other examples of ratios. Ratio does not require a time frame and multiplier. Sex ratio is one example of ratios. It is defined as number of females per 1000 males. It is a mater of great concern that it is declining in India. In 2001 it was 933: 1000.

Dependency Ratio:

Total dependency ratio

$$= \frac{\text{(Population aged 0 to 14 yrs) + (population aged 65 yrs \& above)}}{\text{Population aged 15 to 64 years}} \times 100$$

Child dependency ratio = $\frac{\text{(population aged 0 to 14 yrs)}}{\text{Population aged 15 to 64 years}} \times 100$

Old age dependency ratio = $\frac{\text{(population aged 65 yrs \& above)}}{\text{Population aged 15 to 64 years}} \times 100$

Proportions of persons above 65 years & below 15 years of age are considered to be dependent on the economically productive age group i.e. 15 to 64 years.In India in 1995, it was as below:

Child dependency ratio = 57.8% Old-age dependency ratio = 7.6%
Total dependency ratio = 65.5%

More the dependency ratio more the burden on the economically productive population of a country. This also reflects in the population pyramid of the country.

Proportion -

A proportion is a ratio in which numerator is always included in the denominator. It is usually expressed as percentage. When we say that one must get 50% marks in MBBS examination to get through we are actually telling the proportion or marks out of total marks one must get i.e.

$$\frac{\text{Marks obtained}}{\text{Total marks}} \times 100$$

i.e. $\frac{50}{100} \times 100 = 50\%$

Mortality due to AIDS out of total mortality is another example of a proportion and can be expressed as –

$$\text{Proportional mortality due to AIDS} = \frac{\text{No. of deaths due to AIDS} \times 100}{\text{No. of deaths due to all causes.}}$$

MEASUREMENT OF MORTALITY

Mortality data is the only data that is available in many of the developing countries. Such information is easy to obtain and due to nature of the event probably it is reasonably accurate. Mortality data is the beginning point for many public health studies. Mortality data is useful for studying health trends, for planning, execution and evaluation of various health programmes. They are very much important for epidemiologic research.

There are number of limitations associated with mortality data. These include-

a) All deaths may not be reported. b) The age of deceased and cause of death may not be recorded accurately. c) National and international comparability of mortality data is in question at times due to lack of uniform and standardized methods of collection of data. d) If a single cause of death is recorded, then information regarding associated conditions and risk factors may not be available. e) Coding systems and diagnosing practices may change over time affecting validity of the data. f) Mortality data is almost useless in cases of diseases like mental illness, where there is very low fatality.

Crude Rates:

$$\text{Crude death rate} = \frac{\text{No. of deaths in a year}}{\text{Mid year population of that year}} \times 100$$

$$\text{Natural increase rate} = \frac{\text{No. of lives births in a year - No. of deaths}}{\text{Mid year population}} \times 1000$$

Specific Rates:

$$\text{Infant mortality rate} = \frac{\text{No. of infant deaths in a year}}{\text{No. of live births in that year}} \times 1000$$

$$\text{Neonatal mortality rate} = \frac{\text{No. of deaths within } 1^{st} \text{ 28 days of life in a year}}{\text{No. of live births in that year.}} \times 1000$$

$$\text{Post neonatal mortality rate} = \frac{\text{No. of deaths between 28 days \& } 1^{st} \text{ yr. of life in a year}}{\text{No. of live births in that year.}} \times 1000$$

$$\text{Still birth rate} = \frac{\text{Annual No. of foetal deaths after 28 weeks of gestation}}{\text{Total no. of births in that year.}} \times 1000$$

$$\text{Child death rate} = \frac{\text{Annual No. of deaths between } 1^{st} \text{ \& } 4^{th} \text{ yr. of life}}{\text{No. of live births in that year.}} \times 1000$$

Under five mortality rate (U5 MR) =

$$\frac{\text{Annual no. of deaths under five years of age} \times 1000}{\text{No. of live births in that year.}}$$

Maternal mortality rate =

$$\frac{\text{Annual no.of maternal deaths due to pregnancy \& puerperal conditions} \times 1000}{\text{No. of live births in that yr.}}$$

Prenatal mortality rate =

$$\frac{\text{Annual no. of stillbirths \& deaths in the first 7 days of life} \times 1000}{\text{No. of live births in that year.}}$$

Case fatality rate (Ratio) = $\frac{\text{Total no. of deaths due to a particular disease} \times 100}{\text{Total no. of cases due to the same disease}}$

Proportional mortality rate (Ratio)

a) for a specific disease.

$$= \frac{\text{No. of deaths from the specific disease in a year} \times 100}{\text{Total deaths from all causes in that year.}}$$

b) for a specific age group e.g. under fives

$$= \frac{\text{No. of deaths under 5 years of age in a year} \times 100}{\text{Total no. of deaths in that yr.}}$$

Survival Rate

5 yr. survival rate = $\frac{\text{Total No. of patients alive after 5 years} \times 100}{\text{Total No. of patients diagnosed or treated.}}$

Adjusted or standardized mortality rates: we can get age, sex, religion or race adjusted/standardized death rates by using direct or indirect methods of standardization. Multivariate analysis, regression techniques can also help us in standardization.

STANDARDIZATION OF RATES

There are two methods of standardization: direct and indirect.

Direct standardization

Directly standardized rates give an indication of the number of events, that would occur in a standard population, if the population had the same age specific rates of the local area.

In this method a standard population is selected. It is defined as a population for which the numbers in each age group are known. Combining two different populations can also create standard population.

Then age specific rates of the population whose crude event rate is to be adjusted or standardized is applied to the standard population. Next step is finding out the expected number of events for each age group in the standard population. These are added together for all the age groups to get the total expected events. Last step is to divide the total expected events by the total standard population to get standardized or adjusted rate.

Like age adjusted or standardized rates there also can be sex, religion or race standardized rates.

Let us understand this with following example:

A city has population of 1,00,000 that includes 70,000 females and 30,000 males. Crude death rate for that population is 10 per 1000.

1. Calculate sex specific death rates for that population.

Sex	Mid yr population	Deaths in the year	Sex specific death rate
Males	30000	450	15.0
Females	70000	550	7.85

Now let us take a standard population with 50000 males and 50000 females.
Apply sex specific death rates of above population to the standard population-

Sex	Standard population	Sex specific death rate per 1000	Expected deaths
Males	50000	15.0	750
Females	50000	7.85	392.5

Standardized death rate per 1000 = 750+392.5/100000*1000
= 1142.5/100000*1000=11.42 per 1000

It will be seen that standardization for sex has increased death rate from 10 per 1000 to 11.42 per 1000.

Direct method of standardization is possible only if the actual specific rates in the subgroups (age, sex, religion or race) of the observed population are available. Number of individuals in each subgroup also should be known.

Indirect standardization

In this method age group or sex specific rate of standard population is applied to the age group wise or sex wise distribution of the study population in order to get expected events in the study population.

Indirect standardization does not require area age or sex specific rates but it requires-

1. Age or sex distribution of each group to be standardized
2. Total events in each group to be standardized.
3. Age or sex specific rates for a standard population.

Let us understand this with one example.

In the study population of 100000 there are 60000 Coal miners, 20000 Farmers, 10000 laborers and 10000 others.

Occupation	Number in study population	Occupation specific death rates in standard	Actual deaths	Expected deaths in study population

		population per 1000		
Coal miners	60000	10	500	600
Farmers	20000	12	200	240
Laborers	10000	08	100	80
Other	10000	15	100	150

Total expected deaths = 1070
Actual deaths = 900
Crude death rate for study population = 900/100000*1000 = 9 per 1000
Standardized mortality ratio = Observed deaths/expected deaths * 100
= 900/1070*100 = 84.11

MEASUREMENT OF MORBIDITY

Any subjective or objective departure from a state of physiological well being is termed as morbidity. According to WHO morbidity can be measured in terms of three units-

i) Persons affected.
ii) The illnesses (period or spells of diseases) that affected these persons and
iii) iii) The duration of these illnesses.

The morbidity rates and ratios are useful for following purposes-

a) They help in deciding priorities by describing the extent and nature of disease load in the community.
b) They help in monitoring and evaluation of various activities aimed at control and prevention of the disease.
c) Most of the times they serve as starting point for epidemiological studies aimed at establishing cause effect relationships or to find out suitable strategies for control and prevention of the disease.
d) Contrary to mortality data they provide detailed information about patient and disease in question and hence are more useful for research activities. Frequency, severity and duration of morbidity are measured by morbidity rates and ratios.

Compared to mortality data, morbidity data has following peculiar problems-

i) Death is unique event, illness can occur in spells.
ii) There are various grades of severity in illness.
iii) Death occurs at a point of time but illness exists over a period of time.
iv) Death is precisely defined but illness is not.

Measurement of disease frequency

Commonly used rates are

Incidence rate: It is defined as number of new cases occurring a defined population during a specified period of time.

a) Incidence rate (persons) =

$$= \frac{\text{No. of persons starting an episode of illness in a defined period}}{\text{Average no. of persons exposed to risk in that period}} \times 1000$$

b) Incidence rate (spells) =

$$= \frac{\text{No. of spells of illness starting during the defined period}}{\text{Average no. of persons exposed to risk during that period}} \times 1000$$

c) Attack rate =

= No. of new cases of a specified disease during a specified time interval X 100
Total population at risk during the same interval

d) Secondary attack rate =
= No. of exposed persons developing the disease within
the range of the incubation period X 100
Total no. of exposed/susceptible.

e) Cumulative incidence rate =
= No. of people who get a disease during a specified period X 100
No. of people free of the disease in the population
At risk at the beginning of the period.

Prevalence rate-It refers to all cases (new and old) existing in the community during a specified period.

a) Point prevalence =
= No. of all current cases (old and new) of a specified disease
Existing at a given point in time X 1000
Estimated population at the same point in time.

b) Period prevalence =
= No. of existing cases (old and new) of a specified disease during
a given period of time interval X 1000
Estimated mid interval population at risk.

It will be appropriate to know the relationship between prevalence, incidence and duration of an illness for a stable condition. The relation is expressed as-

Prevalence = Incidence x Duration.

Following are the factors that lead to increase in prevalence of a disease-
1) Longer duration of the disease
2) Prolongation of life of patients without cure
3) Increase in new cases i.e. incidence
4) In migration of cases
5) Out migration of healthy people
6) Improved diagnostic facilities leading to better reporting.

Measurement of duration of illness

Commonly used rates are-

Avg. duration of illness = Sum of duration of illness of cases in a sample
No. of cases in the sample

Days of illness per spell = Sum of days of illness of all sick persons
No. of spells of illness

Measurement of severity

Case fatality rate = No. of deaths ascribed to a specified disease / No. of reported cases of the specified disease.

Disability rates (in part these rates may tell about duration of illness and its complications). There are two types-

a) Event type-e.g. No. of days of restricted activity, bed disability days and work loss days within a specified period.

b) Person type-e.g. limitation of mobility (confined to bed or house) and limitation of activity (limitation to perform the basic activities of daily living).

c) Special measures like Sullivan's index which gives expectation of life free of disabilities. It takes into account comprehensively morbidity and consequences there of that a person can experience throughout his life.

World Health Organization has developed International Classification of Diseases for national and international use. This leads to uniformity in recording diseases as well as facilitates comparability. The latest revision of ICD Published in 1992 classifies diseases into 21 main chapters. There is a hierarchical arrangement of subdivisions (rubrics) within each chapter.

MEASUREMENT OF FERTILITY:

Fertility means the actual bearing of children. It depends upon several factors such as age at marriage, duration of married life, spacing of children, education, economic status, caste and religion, nutrition and use of family planning methods. Following are some rates related to fertility:

Birth Rate = (No. of live births during the year / Estimated mid year population of that year) x 1000

General fertility rate = (No. of live births in an area during the year / Midyear female population of age 15 to 44 (or 49) in the same area in that year.) x 1000

General marital fertility rate = (No. of live births in an area during the year / Midyear female population of age 15 to 44 (or 49) and who are married, in the same area in that year) x 1000

Age specific fertility rate – It is defined as no. of live births in a year to 1000 women in any specified age group.

Age specific marital fertility rate- It is the no. of live births in a year to 1000 married women in any specified age group.

Total fertility rate – This rate gives the approximate magnitude of "completed family size" and is determined by summing the age specific fertility rates for all ages.

Total marital fertility rate – It is the average number of children that would be born to a married woman if she experiences the current fertility pattern through out her reproductive period (15 to 44 or 49 years).

Gross reproduction rate – It is the average number of girls that would be born to a woman if she experiences the current fertility pattern throughout her reproductive span (15 to 44 or 49 yrs.) assuming no mortality.

Net reproduction rate – It is defined as the number of daughters a newborn girl will bear during her lifetime assuming fixed age specific fertility and mortality rates.

Pregnancy rate – It is defined as the rate of number of total pregnancies in a year to married women in the age group of 15 to 44 or 49 years.

Abortion rate – It is given as the number of all types of abortions, usually per 1000 women of child bearing age.

Abortion ratio – It is determined by dividing the number of abortions performed during a particular time period by the number of live births over the same period.

Child-women ratio – It is defined as the number of children of 0 to 4 years age per 1000 women of child bearing age, i.e. 15 to 44 or 49 years of age.

OFFICIAL HEALTH STATISTICS

Official Health Statistics comprises the following broad items of information :

A) Existing Health Situation in India (State of health of the population, Factors influencing the Health of the population)

B) Response of Official Machinery to Improve the Health situation {Services rendered for controlling the factors (preventive) and for treating the health ailments (curative and rehabilitative). Also includes promotive measures}.

Under (A) the following broad items of information may be considered :-

1) Population information
2) Vital statistics
3) Burden of Disease and Disability and health conditions requiring medical attention
4) Cause of death information
5) Reproductive and child health
6) Environmental pollution
7) Availability of basic amenities like safe drinking water, availability of sanitation facilities, source of lighting, housing facility
8) Socio economic condition of family like average family size, level of poverty, literacy level of head of household
9) International comparative information

Under (B) the following broad items of information may be considered :-

10) Health and population policies and intervention strategies
11) Health and medical education infrastructure – both Allopathy and ISM&H
12) Availability and deployment of Health and Medical personnel – both Allopathy and ISM&H
13) Health care delivery organisation and actual infrastructure – both Allopathy and ISM&H
14) National health programmes – both Allopathy and ISM&H
15) Health expenditure and Health care financing
16) Health and medical research
17) Checking food adulteration
18) Preparedness to tackle epidemics and aftermath of natural calamities
19) Availability of specialty treatments and along with cost

HEALTH UNDER THE INDIAN CONSTITUTION:

India has a federal set up with 29 states and 6 union territories. Under the provision of the constitution of India, Public Health is primarily a State subject. Although more effective central action and a larger measure of centre-state cooperation and inter-state coordination is necessary, these do not call for amendment of constitution provisions.

Keeping in view the democratic traditions built-up in this country, the objective can be better achieved by the growth of healthy conventions, greater goodwill and better education, by a system of grants-in-aid by the Central Government in support of public health program of a national character like those of water supply and sanitation, eradication/control of communicable diseases, family planning and schemes of training of health personnel, etc.

The subject matter in the field of Health, divided between Centre and States, as per constitutional provisions are as follows:

Union List:
- Port Quarantine Administration
- Post Graduate Medical Education
- International Health
- Medical Research etc.

State list:
- Public Health
- Sanitation and medical care
- Hospitals and dispensaries

Concurrent List:
- Vital Statistics (including registration of births and Deaths)
- Food Adulteration
- Drug Control etc

Main sources of Health Information are therefore the Central, State Government agencies inclusive of Hospitals, Dispensaries, Community Health Centres, Primary Health Centres and Subcentres and also the Private Sector Health Utilities.

ORGANISATIONAL SETUP:

The organisational setup for the collection, compilation, analysis and interpretation of Health/Related Statistics at Centre:

The details regarding the structure and functions of the Ministry are available in the Annual Report of the Ministry.

The Statistical unit of the Department of Health called Central Bureau of Health Intelligence (CBHI) is a section of the Directorate General of Health Services(DGHS), New Delhi. DGHS is an attached office of the the Ministry of Health & Family Welfare

The Departments of Family Welfare and the Department of ISM&H have separate Statistical Divisions.

At state level, Ministry of Health is headed by Secretary Health and under which there are one or more Directorates (Health/Health & Family Welfare/ Public Health & Preventive Medicine/Family Welfare/Medical Education/Medical and Rural Health Services/Vaidya Vidhana Parishad etc.). In the district, there is a district chief medical and health officer, who is the overall in charge of the public health situation of the district.

HEALTH CARE DELIVERY SYSTEM – GOVERNMENT

At the lowest level of health care delivery sub centres are there for every 5000 general population and 3000 tribal population in rural and tribal areas. This sub centre is manned by 1 multi purpose health worker (male), 1 multi purpose health worker (female) and voluntary worker on honorarium basis. For every 6 sub centres, there is a Primary Health Centre consisting of 15 staff – Medical Officer 1, Pharmacist 1, Nurse Midwife 1, Health Worker Female 1, Health Educator 1, Health Assistant Male 1, Health Assistant Female (Lady Health Visitor) 1, Upper Division Clerk 1, Lower Division Clerk 1, Laboratory Technician 1, Driver 1 (subject to availability of vehicle, class IV 4).

For every 4 PHCs, there is one CHC as a referral centre and it consists of 25 staff – Medical Officers 4, Nurse Midwife 7, Dresser 1, Pharmacist 1, Lab Technician 1, Radiographer 1, Ward Boys 2 etc. The Medical Officers are either qualified or specially trained to work as Surgeon, Obstetrician, Physician, and Pediatrition. One of the Medical Officers should be either qualified for specially trained in public health. Apart from the Community Health Centres, there are TB Hospitals/Sanitoriums, Leprosy Hospitals, Fever Hospitals, ID Hospital, Maternity Home, Municipal Corporation Centres, Rural Family Welfare Centres, Urban Family Welfare Centres, Post Partum Centres, Urban Health post working in Rural/Urban areas. Above the level of Community Health Centres, there are sub divisional/taluk hospitals, Municipal Hospitals, Designated District Hospitals, Other hospitals in district level, General hospitals, Teaching, Non teaching hospitals etc. Specialised Hospitals, state level hospitals are there.

INFORMATION FLOW TO CENTRAL AUTHORTIES

The Directorates of Health Services/Health and Family Welfare Services/Medical Education/ Medical and Rural Health Services are the nodal points at the state level for Central Authorities.

As already explained the rural health facilities such as Sub centre (SC), Primary Health Centre(PHC) and Community Health Centre(CHC) are the grass root level data generation centres. The registers and records at the sub centre contain the basic information about the various activities and the health characteristics of the population, disease incidence etc. The Primary Health Centre is referral unit for a Sub Centre and the Community Health Centre is the referral unit for a Primary Health Centre. As far as information reporting is concerned, the Sub centre reports information to Primary Health Centre and Primary Health Centre and Community Health Centre both report information to District Health

Office(DHO). It may be seen that the information flow is from SCs to PHC and PHC/CHC to DHO to State to Centre.

For the vertically run national health programmes, the nodal officers of the health programmes at district, state and centre receive the information about the programmes (disease).

STATISICAL SET UP AT THE CENTRE AND STATES/UTs

Central Bureau of Health Intelligence is the health statistics wing of Directorate General of Health Services, Deptt. of Health. National AIDS Control Organisation and the vertical national health programmes of Deptt. of Health have their own data compilation arrangement and Central Bureau of Health Intelligence receives data of the programmes/programme disease for their regular annual publication "Health Information of India." The monitoring, evaluation, demographic research and survey unit of Deptt. of Family Welfare is the Family Welfare Statistics Wing. Also, the infrastructure division of Deptt. of Family Welfare collects information on Govt. Rural Health facilities and Health Manpower working in rural areas in the Govt. Health facilities. The Deptt. of AYUSH has a separate statistical unit to collect information on the treatment infrastructure and manpower under AYUSH systems.

At the state level, however, either the State Bureau of Health Intelligence or the state demographer or the State AYUSH officer is responsible for statistical information let into the state. Not all the states have the State Bureau of Health Intelligence; in some States a small statistical unit is there which is mostly involved in getting data regarding the registration of births and deaths or the 100% centrally sponsored Family Welfare programmes.

CENTRAL BUREAU OF HEALTH INTELLIGENCE (CBHI)

As already explained, the Central Bureau of Health Intelligence in Directorate General of Health Services (attached the office of MOHFW), is the Health Statistical Unit of Deptt. of Health.

The CBHI collects information on the following items relating to Health from the primary/ secondary source agencies:-

<u>Education</u>

1. Allopathic Medical/Dental Education from MCI/DCI recognised Medical and Dental Colleges through annual proforma.
2. Paramedical education from DHS/DME of States/UTs to annual proforma

<u>Manpower</u>

3. Registered number of allopathic doctors/dental surgeons/general nurses/pharmacists from the respective All India Councils such as Medical Council, Dental Council and Pharmacy Council, annually
4. Registered number of AYUSH practitioners from Deptt. of AYUSH annually
5. Number of allopathic doctors/dental surgeons working in Govt. establishments in States/UTs from the DHS of States/UTs through annual proforma
6. Number of paramedical personnel working in Govt. establishments in States/UTs from the DHS of States/UTs through annual proforma.
7. Rural Health manpower information from infrastructure division of Deptt. of Family Welfare – annually

Treatment infrastructure

8. Allopathic treatment infrastructure in Govt. sector on annual basis from DHS of States/UTs
9. Rural Health facilities information from infrastructure division of Deptt. of Family Welfare - annually
10. Information on treatment infrastructure from Deptt. of AYUSH – Annually

Public Health Information

11. Reported cases and deaths due to cholera (weekly), communicable (monthly) and non communicable diseases (monthly) through prescribed formats
12. Information on Malaria, Filaria, Japanese Encephalitis, Kala Azar, Dengue from National Vector Borne Disease Control Programme - annually
13. Vertical National Health Programmes information from the Central Programme Officers - annually
14. Information on STD, AIDS from NACO – annually

Cause of death information

15. Information on cause of death from Registrar General of India – annually

In addition to above, CBHI also collects further relevant information from population data to International comparative statistics and brings out regular annual health statistics publication of CBHI titled "Health Information of India".

Health Information of India- Annual Publication of CBHI

The Publication "Health Information of India" (HII) is the regular annual Statistical Publication of CBHI. The combined issue for the years 2000 & 2001 is the latest printed issue.

There are around 130 data tables covering the following topics:
Population & Vital statistics, Per Capita Income, Percent of population below Poverty line, Medical & Dental Education, Registered Personnel Information, Number of Govt Allopathic Hospitals, , No of ISM&H facilities (both Educational & treatment), Production of certain Essential Drugs, Public Health Statistics for cholera & major communicable diseases and few others , Causes of Death Statistics, International Comparative Statistics etc.

The data are presented at State/UT levels in the publication.

The data for the publication are received from the various central agencies such as Census Commissioner , Registrar General Of India (RGI), Central Statistical Organisation (CSO), Planning Commission, the Central Program Officers of Directorate General Of Health Services (DGHS), the Departments of Family Welfare, Indian System of Medicine and Homoeopathy (ISM&H), the various Statutory Councils such as Medical/ Dental/ Pharmacy /Nursing Councils, the Directorate of Health Services (DHS) of States/Uts, etc.
In toto, there are about 40 source agencies(all DHS of States/Uts are clubbed into a single source here) for HII.

The data of Health Information of India are put in CBHI WEBSITE with address www.cbhidghs.org . This Website site has been launched to the public on 9.1.03 by the

Hon'ble Minister of State for Health and Family Welfare Sri A.Raja. This Website contains the data of 1999-2003 publications of HII. Also it contains static pages giving information about CBHI , the in service training programs conducted in Medical Record Science and General and Health Statistics. Provision is there for downloading the application forms for the training programs. Also there is provision for the States/UTs to fill the blank proformae and send to CBHI through E mail.

The CBHI also brings out ad-hoc Publications. "Health Map of India 1998" is one of the ad-hoc publications. This Publication depicts district wise information on Health Facilities in Maps. The Publication 'Directory of Medical, Dental, and Pharmacy Institutions (Educational)' 2001 is the latest ad-hoc publication of CBHI.

Limitations of CBHI Publications:

- Data below State/UT level are not available in central Health Statistical Publications.
- Private Sector Health Data are not included. There is no Statutory Provision to collect Health Data from Private Health establishments.
- Age & Gender specific data are not there. Action is being taken to capture morbidity data, gender wise for the publications.
- Morbidity Surveys are not done regularly and hence, there is a dearth of data on prevalence/incidence rates of diseases.
- There is no systematic generation of health data and scientific methods of recording at grass root levels. As such the response of States/UTs is below the mark in the matter of requests for data from central agencies.
- There are no data on cost of Health Care , availability of specialty treatments in the various Hospitals.
- Hospital Inpatient & Out patient Medical Records are not computerised.
- There is no data on registered Para Medical Professional.
- The registered Health Personnel data are not corrected for death etc.
- There is no District Health Facilities Directory in many of the States/UTs.

NATIONAL AIDS CONTROL ORGANISATION

Through regular sentinel survey and behavioral sentinel survey, data on AIDS cases gender wise, data on HIV prevalence among pregnant women aged 25-49 years, condom use rate among non regular sex partners, STD cases etc. are generated. NACO has a website of its own.

INFRASTRUCTURE DIVISION OF FAMILY WELFARE

Infrastructure Division of Deptt. of Family Welfare collects information on Govt. Rural Health Care Establishments (SC/PHC/CHC) and health/medical manpower in the Rural Health Care Establishments from States/UTs and brings out a 6 monthly bulletin titled Rural Health Statistics in India. March 2002 issue is the latest issue and it contains information as on 31.3.2001. This publication also contains information on training of paramedical personnel.

STATISTICAL DIVISION OF DEPTT. OF FAMILY WELFARE

Statistical Division of Deptt. of Family Welfare brings out the annual publication "Family Welfare Year Book". The latest issue relates to 1997-98. This gives information on Family Welfare programmes such as Immunisation Coverage and Medical Termination of

Pregnancy Services and Family Planning Acceptance, Information on Impact of Family Welfare Programmes such as Characteristics of Acceptors, Information on Outlay and Expenditure on Family Welfare etc. Data are collected from States/UTs and other source agencies.

This unit has also conducted All India National Family Health Survey 1992 and National Family Health Survey 1998; the all India and State Reports are already published. Planning for next survey is going on. Also Rapid Household Survey and District Facility Surveys have been done through International Institute of Population Sciences, Mumbai. The Reports are already brought out.

STATISTICAL DIVISION (PLANNING & EVALUATION CELL) OF DEPTT. OF AYUSH

Statistical Division of Deptt. of AYUSH brings out the annual publication like "AYUSH in India". The latest issue relates to year 2003. This contains information on Medical care facilities manpower, medical education, license to pharmacies, paramedical to courses, outlay and expenditure, Drug testing laboratories, qualitative information such as diseases for which patients generally visit AYUSH hospitals, dispensaries etc. The sources of information are the different States/UT governments, the different councils under AYUSH etc.

HEALTH DATA SOURCES:

Health Data are mostly collected through complete enumeration.

Complete Enumeration

Central Agencies:

Ministry of Health & Family Welfare (MOHFW):

Sample Survey

INTERNATIONAL STATISTICAL CLASSIFICATION OF DISEASES AND RELATED HEALTH PROBLEMS

Hospitals Medical Records are a vast source of information and for coding Medical Records (Inpatient & Outpatient records), the International Statistical Classification of Diseases and Related Health Problems (ICD 10) is used. ICD 10 introduced in 1993 by the World Health Organisation has three volumes. Volume I is the detailed classification. Volume II is the manual of instructions for use of Vol. I and Vol. III. Volume III is the alphabetical index. ICD 10 is an alphanumeric classification. ICD 10 is adopted by India in the year 2000. CBHI is conducting one week short term orientation training on ICD 10 for Inservice candidates of Govt. and Non Govt. sector through its training centres RHSTC Mohali, Punjab and Regional Office of Health & Family Welfare, Bangalore, Karnataka.

HOSPITAL (MEDICAL STATISTICS)

We know that statistics is the science of collecting, classifying, summarizing, analyzing and interpreting data. People concerned with the administration and management of a hospital should find out statistical data related to services rendered by the professional departments. The primary needs for hospital statistics are-

i) To establish administrative control over functional activities. ii) To provide reports to the governing board, outside agencies etc. iii) To provide a basis for preparing operating budgets. iv) To provide a basis for the distribution of expenses when computing cost of operation. v) To provide a basis for the calculation of average income and costs per unit of

service rendered. vi) Realistic planning for the future is impossible without basic statistics. vii) To assess utilization of hospital facilities. viii) To provide data for health intelligence to public health authorities.

Hospital management requires statistical data which will provide quantitative information concerning the scope of activities within the hospital. It helps the administration to ascertain in increase or decrease in the volumes of work and the effective utilization of men and material.

Following is a list of definitions and rates commonly used in hospital statistics-

Bed Complement – The bed complement of a hospital is the total number of hospital beds normally available for use by inpatients.

Patients Day – For statistical purposes the hospital census of the total number of inpatients in a hospital on any day is counted at 23.59 hours. A patient day is therefore a period of service rendered to an inpatient between the census taking hours on two successive days, the day of discharge being excluded.

Average Daily Census – It is the average number of in patients excluding newborns receiving care each day throughout a given period of time. It is calculated by summing up the daily census of that period and dividing it by the number of calender days in that period.

Percentage of Occupancy – Percentage of occupancy or occupation rate is the ratio of actual patient days to the maximum possible patient days (based on bed complement) during any given period of time.

Average Length of stay – It is the average number of patient days of service rendered to each inpatient during a given period.

Turnover Interval – It is the average period in days a medical bed remain vacant between one discharge and another admission.

Throughput – The throughput is the number of separations (discharges, deaths and transfers) divided by the mean number of available beds.

Death Rates- The gross death rate is the percentage of total number of hospital deaths during any given period in relation to total discharges and deaths during the period. Some important death rates are-

a) Anaesthetic death rate
b) Post operative death rate
c) Maternal death rate
d) Neonatal death rate.

Educational Statistics in India

Introduction

India's commitment to the spread of knowledge and freedom of thought among its citizens is reflected in its Constitution. The Directive Principle contained in Article 45 enjoins that the State shall endeavour to provide free and compulsory education for all children until they complete the age of fourteen years. Special care of the economic and educational interests of the underprivileged sections, particularly, the Scheduled Castes and Scheduled Tribes is laid down as an obligation of the State under article 46. Though education is in the concurrent list of the Constitution, the State Governments play a very major role in the development of education particularly in the primary and the secondary education sectors.

Broadly speaking, the Central Government is responsible only to plan, to coordinate and to guide educational activities in the country. Statistics being the basis for planning, it is therefore vital to have systematic coordination and to provide guidance for effective discharge of the Central responsibility in this field. In order to discharge its responsibility adequately, the Ministry of Human Resource Development (Deptt. Of Secondary & Higher Education) collects the necessary educational statistics from the country in the prescribed forms. The main sources of this information are the Education Departments of the State Governments/ U.T. Administrations and the Secondary/Senior Secondary Boards/Universities.

From the State Governments and U.T. Administrations, the statistics are being collected in ES series of Forms i.e. ES.I(S) to ES.IV (S) for school level educational data and ES.I (C)to ES.IV (C), for higher level educational data along with proforma for selected educational statistics and distance education statistics. These forms cover the entire educational system in the country, right from the pre-primary to the post-graduate and doctorate level including distance education. Moreover, it is a record of a total count, inter-alia, of all the recognized educational institutions, students, teachers, education expenditure, output etc.

School level ES forms:
ES-I(S): Numerical data (classwise and coursewise enrolment and teachers)
ES-II(S): Financial data
ES-III(S): Examination results
ES-IV(S): Data in respect of SC/ST
Higher level ES forms:
ES-I(C): Numerical data (classwise and coursewise enrolment and teachers)
ES-II(C): Financial data
ES-III(C): Examination results
ES-IV(C): Data in respect of SC/ST

Statistical Organisation in States :- The educational data pass through a long channel starting from the educational institutions to the national level. The various educational authorities in between handle it according to their responsibilities. All States/ UTs have a statistical unit in the education department at its headquarters. They collect a part of the data direct from the institutions, while for the rest they have to depend upon the district authorities. At the district level also, a part of the data is collected direct from the institutions and for the rest it has to depend on the Block/Taluka office. The State office, however,

collects the entire higher education data from the field without the help of any intermediary agency.

A Model Procedure:-In order to reduce time lag at each level a model procedure consisting of four stages of the work of collection of educational statistics is listed below which may be followed by all States/UTs.

Stage-1: Preparation, printing and distribution of institutional returns to educational institutions:

i) Every year, the Ministry of Human Resource Development sends blank copies of the E.S. series forms to all States/UTs for the supply of the requisite statistics at school and higher education levels. The first task before the State governments is to prepare institutional returns for school and higher education separately in the light of ES series of forms. The additional data items required by States/UTs can also be incorporated in these returns. The specified finalized Institutional Returns are made available to all the institutions located in their State/UT before 30th September of the academic year.

ii) The language of the returns is another point that has to be kept in mind at this stage. There are institutions where the teachers may not be conversant with English; it is necessary that the returns for such institutions are printed in the regional language.

iii) The statistical unit at the State Headquarters is a nodal agency for distributing returns to universities, colleges and other institutions of higher education and also to District Education Offices for onward distribution to High & Higher Secondary schools, teacher training schools and tehsil/taluk office for distribution to primary and middle schools. In short, the authority who is responsible for the collection of institutional returns from a particular type of institutions should normally supply the blank copies of the relevant return. Before the distribution of blank copies of the return to the educational institutions, it is essential that the concerned authorities at the state, district and tehsil/taluk level should have a complete and up-to-date list of all the institutions from which they have to collect statistics. This list should be maintained and updated regularly. Further, it will be useful to maintain information about the courses available in each of these institutions, the minimum admission requirements for each course and the duration thereof.

Stage-2 : Collection of institutional returns :

This is the most crucial stage in the entire scheme of this work. It appears that the maximum delay occurs at this level. In spite of this, it is not known whether the coverage is always cent percent complete or not. It is, therefore, extremely important to be vigilant so that no educational institution remains uncovered and all of them must send their returns to the concerned educational authorities within predetermined time schedule. The following suggestions may be followed :

(i) The collection of returns from the departmental institutions normally does not present any difficulty and if there is any, the same may be resolved by taking up the issue at an appropriate level. Morever, the Education Department should issue necessary instructions to the tehsil/taluk and district officers and the head of Government institutions for timely submission of the statistical return to the appropriate authority.

(ii) In regard to institutions under the administrative control of State Departments (other than the Education departments), the State Government may be approached to issue instructions to all the concerned departments to cooperate fully with the Education

Department in the matter of collection of statistics. This also applies to institutions under local bodies including Zilla Parishads and Panchayat Samities.

(iii) The necessary support may be obtained from universities in case of delay in getting statistics from any college. In States, where University Grants Commission is functioning, their cooperation can prove valuable in this task.

(iv) A large majority of private institutions receive aid from Government. If the grant-in-aid rules include a specific clause for the submission of timely and accurate return by the institution to the concerned authorities it can go a long way in ensuring complete coverage of all such institutions in the educational statistics of the State.

(v) This leaves behind a small element of unaided institutions. It is hoped that these will also respond favourably to the request of the Education Department who grants recognition to them.

(vi) The system of school inspection can also help a good deal in this work particularly in relation to private institutions. The inspection form should contain a specific provision for reporting whether the institution maintains the necessary records for educational statistics and whether it has been submitting them regularly. A list of the necessary record to be maintained in this connection may, however, be circulated to all the institutions well in advance.

(vii) The heads of all institutions may be requested to make a member of staff (preferably the Mathematics teacher) responsible for accurate compilation of the data in the institutional return and its timely submission to the authorities concerned. This special duty should be taken into account in assessing the worth of the teacher.

(viii) In certain states, periodic meetings of teachers are held in which, inter-alia, school returns are collected. It will be useful to exploit this occasion to the maximum advantage for statistical purposes.

ix) The primary and middle schools should send their returns to the tehsil/taluk office from where these may be forwarded to the district office. The district office, in addition should be responsible for the collection of returns from high/higher secondary schools, teacher training schools, etc. Apart from receiving district returns, the State Headquarters should collect returns from colleges and institutions under other Departments of the State Government. It may be noticed from the above that the offices at the State and District levels are the two most vital links in the system of collection of data.

Stage-3 : Scrutiny of returns and the reconciliation of discrepancies:

For the accuracy of the data, the returns must be scrutinized thoroughly and discrepancies therein if any, reconciled before it is consolidated. It will be better if the reconciliation is done at the initial stage. The levels at which the institutional returns should be scrutinized are as follows :-

The scrutiny of data should start as soon as the very first return is received in tehsil/taluk/District/State office. The tehsil/talnk office may be concerned with the scrutiny of returns of all the primary and middle schools in the area. After reconciliation of discrepancies, all the returns should be submitted to the District office for consolidation. The District office should scrutinize these returns in addition to the returns received from high/higher secondary schools, teacher training schools before consolidation/feeding in computer. Similarly, the office at the state level should scrutinize all the institutional returns it gets as well as those received from subordinate and other office before taking up feeding in computer consolidation. It will be necessary to complete this work as quickly as possible and necessary clarifications/revision of data may be obtained from concerned institutions/offices.

The follow up work in regard to the reconciliation of discrepancies at each level is as important as that of collection of institutional returns. This work should also be pursued with the same vigor and urgency as is suggested for the collection of institutional returns. The occasional meetings of school teachers can also be utilized with advantage for this purpose. At the State level, the practice of calling dealing hands from the subordinate offices for the reconciliation of discrepancies is yet another step which helps a great deal and may be used, where necessary.

Step-4 : Tabulation and consolidation of data and preparation of E.S. series of Forms

The data of primary and middle schools may be consolidated at three stages i.e. tehsil/taluk, District and State levels. The consolidation of statistics of high/ higher secondary schools, teacher's training schools, etc., may be done at the district level and that of the remaining institutions at the State level. Consolidation of data should be done in consolidation registers/Computer file. As soon as any return is found to be finally or even fairly accurate, the posting of data in consolidation registers/Computer file should start at once. It is also not necessary to wait till all the returns are received before starting the consolidation work. The officer-in charge of statistical work at the tehsil/taluk office, district office and at the State headquarters should keep the progress of collection, scrutiny, reconciliation and consolidation work up to date and inform the same to the concerned higher office fortnightly. The consolidated data may then be posted in the prescribed forms very carefully.

CURRENT STATUS OF EDUCATIONAL STATISTICS

The major responsibility of collection, compilation and dissemination of educational statistics from institutions lies with the MHRD and other agencies, like NCERT, UGC, AICTE, etc. Household data collected by RGI and NSSO also contain information relating to education and allied areas. The major coverage by each such agency is presented below:

Sl. No.	Agency	Major Coverage			Primary Source	eriodicity	Data Collection Mode
A	MHRD	i)	Institution, enrolment, examination, teachers for formal education covering all the stages, forms and types of Education		Institution	Annual	Census
		ii)	Distance Education			Annual	
		iii)	Sourcewise and objectwise expenditure of States/Uts			Annual	
B	NCERT	a)	Schools Education		School		
			i)	Schooling facility in rural areas	Institution	About 8 years	Census and Sample

			ii)	Schools, enrolment, teachers, educational inputs, physical, facilities, basic amenities, incentive schemes			
			iii)	Courses of studies, institution, enrolment at +2stage; enrolment and teachers for non-formal education, special education and pre-primary education			
		b)	Teacher Education Coursewise enrolment, Teachers' qualification etc.				
			i)	Elementary Teacher Education	Institution	Ad hoc	Census
			ii)	Secondary Teacher Education	Institution	Ad hoc	Census
			iii)	DIETs	Institution	Ad hoc	Census
C	UGC	i)	Institutions		Institution	Annual	Census
		ii)	Enrolment coursewise teachers for higher education				
D	Association of Indian University	i)	Institution, subject, course, finance, etc.		Institution	Annual	Census
		ii)	Admission procedures number of seats, etc. About courses on higher education.		Institution		Census
E	AICTE	Intake capacity of technical education institutions			Institution	Annual	Census
F	IAMR	i)	Manpower profile disciplinewise		Institution		Yearly
		ii)	Institution, intake and outtake data through NTMLS for engineering				

G	Director -ate of Adult Educatio n	Adult Literacy enrolment, targets expenditure	Project	Annual	Project Profile
H	Planning Commi- ssion	Data on MNP	States/Uts	Annual/five Year Plan	
I	RGI	Educational facility, and literacy by levels of education. Etc.	Household	Decennial	Census
J	NSSO	Enrolment, drop-out and private expenditure	Household	Quin- quenrial	Sample
K	NIEPA	Educational Administration	State Govt.	Adhoc	Census
L	DPEP	School at Primary level	Project School	Annual	Census

In addition to the above, there are other organisation which provide data on education. These are given below.

i) Directorate of Employment & Training for data on Unemployment collected through Employment Exchange.

ii) Indian Council of Medical Research for information on medical education in India.

iii) Director General of Health Services, Ministry of Health and Family Welfare collect data on health , manpower and vital statistics in the country.

iv) National Family & Health Welfare Survey, 1992-93.

Ministry of Shipping Road Transport and Highways - Transport Research Wing (TRW)

The TRW in the Ministry of Shipping, Road Transport and Highways is the nodal agency for providing research inputs, data support and analysis for the various wings of the Ministry; thus enabling policy planning, coordination and evaluation of the performance of the transport modes. The TRW is responsible for the collection, compilation and dissemination of data relating to roads, road transport, ports, shipping, ship building, ship repairing and inland water transport. In this context, the TRW brings out annual publications on Roads and Road Transport, Ports, Shipping and Inland Water Transport. In addition, the Wing also brings out performance review of public sector passenger road transport operations. The inputs from TRW also go towards estimation of GDP for road and water transport sectors.

India's transport system comprises of a number of distinct modes and services. These include:

(i). Railways
(ii) Roads
(iii) Road transport
(iv) Ports
(v) Inland Water Transport (IWT)
(vi) Coastal shipping
(vii) Airports and
(viii) Airlines

Crime Statistics

Basics of Crimes

The Criminal Procedure Code (Cr.P.C.) divides all the crimes into two categories.
(i) Cognizable (Sec.2(c) CrPC).
(ii) Non-cognizable (Sec.2(1) CrPC).

Cognizable Crimes

All cognizable crimes reported in the country are dealt by the Police in which a police officer may arrest a person without a warrant. In such crimes, the police have a direct responsibility to take immediate action on receipt of a complaint or of credible information, visit the scene of the crime, investigate the facts, apprehend the offender and arraign him before a court of law having jurisdiction over the matter. Cognizable crimes are broadly categorised as those falling either under the 'Indian Penal Code (IPC)' or under the 'Special and Local Laws (SLL)'.

Cases under SLL represent preventive policing i.e. reported crime indicate better policing efforts. The Central Acts are enactments on special subjects which are applicable in the whole country while Local Acts are applicable to a particular part of India.

Non-Cognizable Offence

Non-Cognizable crimes are generally left to be pursued by the affected parties themselves in Courts. Police does not initiate investigation in non-cognizable crimes except with magisterial permission. The first schedule of the Cr.P.C. gives the classification of the offences of the IPC into cognizable & Non-cognizable.

The various crimes, that are being recorded, can be broadly grouped under the following categories for Statistical Information System.

Broad classification of crimes under the Indian Penal Code (IPC)

i) **Crimes Against Body**: Murder, Attempt to Murder, Culpable Homicide not amounting to Murder, Kidnapping & Abduction, Hurt, Causing Death by Negligence;
ii) **Crimes Against Property**: Dacoity, Preparation & Assembly fro Dacoity, Robbery, Burglary, Theft;
iii) **Crimes Against Public Order**: Riots, Arson;
iv) **Economic Crimes**: Criminal Breach of Trust, Cheating, Counterfeiting;
v) **Crimes Against Women**: Rape, Dowry Death, Cruelty by Husband and Relatives, Molestation, Sexual Harassment and Importation of Girls;
vi) **Crimes Against Children**: Child Rape, Kidnapping & Abduction of Children, Procuration of minor girls, Selling/Buying of girls for Prostitution, Abetment of Suicide, Exposure and Abandonment, Infanticide, Foeticide;
vii) Other IPC crimes.

Crimes under the Special and Local Laws (SLL)

Special Laws

i) Arms Act
ii) N.D.P.S. Act
iii) Gambling Act
iv) Excise Act

v) Prohibition Act
vi) Explosives & Explosive Substances Act
vii) Immoral Traffic (Prevention) Act
viii) Indian Railways Act
ix) Registration of Foreigners Act
x) Protection of Civil Rights Act
xi) Indian Passport Act
xii) Essential Commodities Act
xiii) Terrorist & Disruptive Activities Act
xiv) Antiquity & Art Treasure Act
xv) Dowry Prohibition Act
xvi) Child Marriage Restraint Act
xvii) Indecent Representation of Women (P) Act
xiii) Copyright Act
xix) Sati Prevention Act
xx) SC/ST (Prevention of Atrocities) Act
xxi) Forest Act
xxii) Other crimes (not specified above) under Special and Local Laws including **Cyber Laws** under Information Technology Act (IT), 2000

Planning in India

What is Planning?

Planning is a processes to make an assessment of all material, capital and human resources of the country including technical personnel and to explore the possibilities to augment them. With a view to have desired long and medium term outcomes in terms of growth and equity and prioritization of application of these resources in various sectors. The participation of communities/ stake holders is a must so that they also have a feeling of belongings to the plan.

Evolution of Planning process in Independent India

The task of rebuilding the country which was partitioned and had devastating impact of second World War was the prime agenda of the Government after achieving Independence from British rulers on 15th August, 1947. The proclamation of the Industrial Policy of 1948, was the outcome of the stupendous work of the Advisory Planning Board. The 1948 resolution proposed the formation of a National Planning Commission to formulate programmes of development and to secure their execution.

ROLE, COMPOSITION AND FUNCTIONS

The Planning Commission was constituted in March, 1950 by a Resolution of the Government of India, and it works under the overall guidance of the National Development Council. The Planning Commission consults the Central Ministries and the State Governments while formulating Five Year Plans and Annual Plans and also oversees their implementation. The Commission also functions as an advisory Planning body at the apex level.

FUNCTIONS

The following functions have been assigned to the Planning Commission as per Government of India (Allocation of Business) Rules, 1961:

(a) Assessment of the material, capital and human resources of the country, including technical personnel, and formulation of proposals for augmenting such of these resources as are found to be deficient.
(b) Formulation of Plan for the most effective and balanced utilization of the country's resources.
(c) Definition of stages in which the Plan should be carried out on a determination of priorities and allocation of resources for completion of each stage.
(d) Determination of the nature of machinery necessary for the implementation of the Plan in all its aspects.
(e) Identifying the factors which are tending to retard economic development and determine the conditions which in view of current social and political situation should be establishment for the successful execution of the Plan.
(f) Appraise from time to time the progress achieved in the execution of each stage of the Plan and recommend adjustment of policies and measures that such appraisal may show to be necessary.
(g) Public Co-operation in National Development.
(h) Specific programmes for area development notified from time to time.
(i) Perspective Planning.
(j) Institute of Applied Manpower Research.
(k) The overall coordination of the Pradhan Mantri Gramodya Yojana.

Basics of Price Index Numbers

Changes in prices influence all economic activities. Constant watch on prices is thus necessary for monitoring and regulation of economic policies. Knowledge of price movement is required for planning and policy making. The changes in prices are measured by the statistical device of index number.

Methods of Constructing Price Index Numbers:

The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by consumers for a market basket of consumer goods and services.

Simple Aggregative Method

In this method, the aggregate price of various commodities in a given year is expressed as a percentage of the same in the base year. The formula is

$$P_{01}=\Sigma p_1/\Sigma p_0 * 100$$

Where Σp_1= aggregate of prices for the current period.

And Σp_0= aggregate of prices for the base period.

This method is simple but has following limitations:-

(i). The prices of various commodities may be quoted in different units, e.g., cereals may be quoted in Rs. Per quintal, liquids like milk, petrol, kerosene may be quoted in Rs. Per litre; cloth may be quoted in Rs. Per metre and so on. Thus the index is influenced very much by the units in which commodities are quoted and accordingly some of the commodities may get more importance because they are quoted in a particular unit.

(ii) In this method the various commodities are weighted according to the magnitudes of their prices and accordingly commodities which are highly priced exert a greater influence on the value of the index than the commodities which are low priced.

(iii)The relative importance in use of various commodities is not taken into consideration.

Weighted Aggregative Method:

In this method appropriate weights are assigned to various commodities to reflect their relative importance in the group. Usually the quantities consumed, sold or marketed in the base year or in a given year are used as weights. If W is the weight attached to a commodity then the weighted price index is given by

$$P_{01}=(\Sigma p_1 W/\Sigma p_0 W) * 100$$

By the use of different types of weight, a number of formulae have emerged for the construction of index numbers.

(a) Laspeyres' Method: Laspeyres Index is obtained by replacing W in the above equation by q_0 the base year quantity which is given below:-

$$^{La}P_{01}=\Sigma p_1 q_0/\Sigma p_0 q_0 * 100$$

(b) Paasche's Method: Paasche's Index is obtained by replacing W in the above equation (1) by q_1 the current year quantity which is given below:-

$$^{Pa}P_{01}=\Sigma p_1q_1/\Sigma p_0q_1 * 100$$

(c) Fisher's Method: Fisher Price Index Number is given by the geometric mean of Laspeyres and Paasche's formulae. Symbolically,

$$^{F}P_{01}=(^{La}P_{01} * ^{Pa}P_{01})^{1/2} =(\Sigma p_1q_0/\Sigma p_0q_0 * \Sigma p_1q_1/\Sigma p_0q_1)^{1/2} *100$$

An interesting property of Laspeyres method is that it is generally expected to overestimate or to leave an upward bias, but the Paasche's method to underestimate, i.e., show a downward bias. This is because in case of price increase there is usually a reduction in the consumption of those items for which the increase has been the most pronounced and hence, by using base year quantities we will be giving too much weight to the prices that have increased the most and the numerator of the Laspeyres index will be too large. When the prices go down, consumers often shift their preference to those items which have declined the most and, hence, by using base period weights in the numerator of the Laspeyres index we shall not be giving sufficient weight to the prices that have gone down the most and the numerator will again be too large. Similarly, because people tend to spend less on goods when their prices are rising the use of the Paasche's or current weighting produces an index which tends to underestimate the rise in prices, i.e., it has a downward bias. But the above arguments do not imply that Laspeyres index must necessarily be larger than the Paasche's.

Fisher's Price Index number is known as 'Ideal' due to the following reasons:-

(i) It is free from bias, since the upward bias of Laspeyres' index number is balanced to a great extent by the downward bias of Paasche's index number.

(ii) It is based on the geometric mean, theoretically which is considered to be the best average for constructing index numbers.

(iii) It conforms to certain tests of consistency.

(iv) This formula takes into account the influence of the current as well as the base year.

The Criteria of a Good Index Number

Mathematical Tests. The components of errors in the construction of index numbers can be broadly classified as:

(a) Formula Error, (b) Sampling Error, and (c) Homogeneity Error.

Formula error arises due to the usage of different formulae, none of which measures the price changes or quantity changes with perfection or exactitude. Sampling error results from the sampling of the commodities to be included in the index for measuring the 'price changes' or 'quantity changes'. Change in the composition of commodities in the two periods of comparison gives rise to homogeneity error.

As a measure of the formula error a number of mathematical tests, discussed below, have been suggested.

(a) Unit Test. This requires the index numbers to be independent of the units in which the prices and quantities of various commodities are quoted. This test is satisfied by all the formulae except the simple aggregate method.

(b) Time Reversal Test. This is one of the two very important tests proposed by Irving Fisher as tests of consistency for a good index number. The test is that the formula for calculating an index number should be such that it will give the same ratio between one point of comparison and the other, no matter which of the two is taken as base. Or, putting it another way, the index number reckoned forward should be the reciprocal of that reckoned backward, except for a constant of proportionality.

Thus, if the time script of any index (say, Price) formula be interchanged then the resulting index should be the reciprocal of the original index.

The Laspeyre's and Paasche's formula do not satisfy Time Reversal Test. The Fisher's ideal index satisfies Time Reversal Test.

(c) Factor Reversal Test. This is the second test of consistency suggested by Irving Fisher. Just as our formula should permit the interchange of two times without giving inconsistent results, so it ought to permit interchanging the prices and quantities without giving inconsistent results – i.e., the two results multiplied together should give the true value ratio, except for a constant of proportionality.

The Fisher's ideal index satisfies Factor Reversal Test. It may be pointed out that none of other formulae satisfies the factor reversal test. Since Fisher's index satisfies both Time Reversal and Factor Reversal Tests, it is termed as 'ideal' index number.

Practical Exercise

On the basis of the following information, calculate the Laspeyres', Paasche's and Fisher's Index Numbers.

Commodity	Base Year		Current Year	
	Price	Quantity	Price	Quantity
A	2	40	6	50
B	4	50	8	40
C	6	20	9	30
D	8	10	6	20
E	10	10	5	20

Consumer Price indices compiled at national level

There are four Consumer Price Indices (CPI) released at national level. These are CPI for Industrial Workers (IW), CPI for Agricultural Labourers (AL), CPI for Rural Labourers (RL) and CPI for Urban Non-Manual Employees (UNME). While the first three are compiled and released by the Labour Bureau in the Ministry of Labour, the fourth by the Central Statistical Organization (CSO) in the Ministry of Statistics and Programme Implementation.

Consumer Price Index for Industrial Workers

Consumer Price Index for Industrial Workers [CPI(IW)] is compiled by the Labour Bureau, Ministry of Labour.

Target Population:

A working class family was defined as one (i) which was located within the center, (ii) which had at least one member working as manual worker in an establishment in any of the seven sectors of employment covered viz., factories, plantations, mining, ports and docks, public motor transport undertakings, electricity generating and distributing establishments, and railways; and (iii) which derived 50 per cent or more of its income during the calendar month preceding the day of enquiry through any manual work.

Source of weight:

The present series of CPI (IW) is on base 2001=100. The weighting diagrams for the purpose of compilation of index numbers had been derived on the basis of average monthly family expenditure of the working class as obtained from the Working Class Family Income Expenditure Survey conducted during 1999-2000.

Price Collection mechanism:

For CPI (IW), the retail prices in respect of selected items of goods and services are collected from 78 centers by the officials of various state Directorates of Economics and Statistics or State Labour Commissioners, etc. The number of items in the consumption baskets of different centers generally varies from center to center, depending upon the prevailing situation in each center and the consumption pattern of the center. The various items of goods and services are classified into six groups namely, (i) food, (ii) pan, supari, tobacco and intoxicants, (iii) fuel and light, (iv) housing, (v) clothing, bedding, and footwear, (vi) miscellaneous. The prices of items like cereals, vegetables, oils and fats, etc., which change frequently, are collected on weekly basis, and the prices of items such as clothing, furniture, utensils, household appliances, durable goods etc. are collected on monthly basis, from two selected outlets/shops in the selected markets. The data on prices of commodities which are supplied through subsidized outlets (fair price shops), and their availability in these outlets are also collected so that weighted average prices can be worked out.

Release of Index:

CPI (IW) is released every month (on the last working day of the following month) for each of the selected 78 centers as well at all India level. It is mainly used for the determination of dearness allowance being paid to Central/State Government employees as also to the workers in the industrial sectors besides fixation and revision of minimum wages in scheduled employments. It is also used in moving the base year poverty lines for urban areas to the subsequent years.

Consumer Price Index for Agricultural Labourers and Rural Labourers

Consumer Price Index for Agricultural Labourers and Rural Labourers {CPI (AL/RL)} is compiled and released by the Labour Bureau, Ministry of Labour.

Target Population:

A rural labour household is defined as one, which derives major income during the last 365 days from wage, paid manual employment (rural labour) vis-à-vis wage paid non-manual employment as also self employment. From amongst the rural labour households, those households, which earn 50% or more of their total income (from gainful occupation) during the last 365 days from wage paid manual labour in agriculture, are categorized as agricultural labour households. Agriculture labour households is a subset of rural labour households. A person is considered engaged in agricultural labour if he/she follows one or more of the agricultural occupations in the capacity of wage paid manual labourer, whether paid in cash or kind (excluding exchange labour) or both. A person who does manual work in return for wages in cash or kind or partly in cash and partly in kind (excluding exchange labour) is a wage paid manual labourer. Persons who are self-employed doing manual work are not treated as wage paid manual labourers. People living in rural areas and engaged in manual labour by working in agricultural and/or non-agricultural occupations in return for wages paid either in cash or kind (excluding exchange labour) or both, are considered as rural labourers. Thus, rural labourers include both agricultural labourers and other labourers.

Sources of weight:

The present series of CPI (AL/RL) is on base 1986-87=100. Estimates of consumer expenditure generated from the results of NSS 38th round (1983), formed the source of weights for the different items of goods and services, used in compilation of CPI (AL) and CPI (RL).

Price collection mechanism:

For both the current series of CPI (AL) and CPI (RL), the retail prices in respect goods and services are collected on monthly basis, from fixed markets in 600 sample villages in 20 states by the National Sample Survey Organisation. The various items of goods and services are categorized into four main groups namely, (i) Food, (ii) Fuel and Light, (iii) Clothing, bedding and Footwear; and (iv) Miscellaneous. The items of goods and services are common but the varieties of most of the items differ from village to village. The collection of retail prices is staggered over four weeks of a month with one-fourth of the sample covered every week.

Release of Index:

CPI (AL/RL) is released every month (during the last week of the following month) for each of the selected 20 states as well at all India level. It is mainly used for the determination of fixation and revision of minimum wages in agricultural sector. CPI(AL) is also used in moving the base year poverty lines for rural areas to the subsequent years.

Consumer Price Index for Urban Non-Manual Employees

Consumer Price Index for Urban Non-Manual Employees [CPI (UNME)] is compiled and released by Central Statistical Organisation, Ministry of Statistics and Programme Implementation.

Target Population:

An urban non-manual family was defined as one, which derived 50 per cent or more of its income from gainful employment on occupations of one or more of its members doing non-manual work in the non-agricultural sector in the urban areas.

Source of weight:

Present series of CPI (UNME) is on base 1984-85=100. The weighting diagram for the current series of CPI (UNME) was prepared on the basis of the data collected through family living survey for urban non-manual employees conducted during 1982-83 in 59 urban centers. The centers were selected keeping in view the (i) concentration of UNME population in the center, (ii) inclusion of state capital cities and (iii) regional representation.

Price collection mechanism:

For the current series of CPI (UNME), 1022 quotations of retail prices in respect of selected items of goods and services are collected on monthly basis, from selected markets in 59 urban centers by National Sample Survey Organisation (NSSO). The number of items of goods and services, for which retail prices are collected for compilation of CPI (UNME), varies from center to center. The various items of goods and services are grouped into five main groups, namely, (i) food, beverages and tobacco, (ii) fuel and light, (iii) housing, (iv) clothing, bedding and footwear; and (v) miscellaneous.

Release of index:

CPI (UNME) is released every month (25^{th} of the following month) for each of the selected 59 urban centers and also at all India level.

Construction of Index Numbers of Wholesale Prices in India- Concepts, Compilation Procedure and related issues.

Concept of Wholesale Prices:

Wholesale Prices represent bulk transactions generally at the early stage of trading. The prices pertaining to bulk transactions may have divergent nomenclatures e.g., farm harvest prices, wholesale prices, procurement prices, support prices, administered prices, import prices, forward prices etc. According to the Office of the Economic Adviser (OEA), Ministry of Commerce and Industry, which compiles the Wholesale Price Index, wholesale prices relate to the transactions at the primary stage which broadly correspond to producer prices.

Need for Wholesale Price Index Numbers:

The Changes in prices, both absolute and relative, influence a wide range of economic activities, and a constant watch on prices becomes necessary for the operation of regulation of current economic policies as well as for the purposes of planning and policy formulation. The changes in prices over a period of time can be gauged by the statistical device of index number of prices. The price index can be either at the wholesale level or at the level of the retail end of the marketing channel. The Wholesale Price Index is important, as it helps in understanding the movement of prices relating to bulk transactions of purchases, usually for further sale. Conventionally, the wholesale price index has been used to measure the variations in exchange value or purchasing power of money. It has also been utilized for other purposes like forecasting changes in business conditions indicating supply-demand relationships, deflating aggregates such as national income, etc.

History of Wholesale Price Index Numbers in India:

The Office of the Economic Adviser to the Government of India undertook to publish for the first time, an index number of wholesale prices, with base week ended August 19, 1939=100, from the week commencing January 10, 1942. The index was calculated as the geometric mean of the price relatives of 23 commodities classified into four groups: 1) food & tobacco 2) agricultural commodities 3) raw materials, and 4) manufactured articles. Since 1947 the index is being published regularly. The WPI series was revised from time to time with new bases viz 1952-53(1948-49), 1961-62, 1970-71,1981-82 and 1993-94. The current series is with 1993-94 base and has total of 435 items.

Choice of the Base year:

The base year of the price index series has to be carefully chosen since it has a considerable influence on the movement of price relatives of individual commodities and on the weighting pattern, thus, influencing the movement of the index as a whole. The Base Year being chosen should be (i) a normal year i.e, a year in which there are no abnormalities in the levels of production and trade and in the price levels and price variations, (ii) a year for which reliable price and other required data are available and (iii) a year as recent as possible. In accordance with these criteria, after examining the data relating to production in different sectors of the economy as well as data on prices and several other related indices, the working group decided to take 1993-94 as the base year for the new series.

Selection of items, varieties/grades, markets and sources of price data:

With a view to reflecting adequately the changes that have taken place in the structure of the economy, almost all the important items being transacted in the economy have been included in the current basket of commodities. The selection of items for the new series was based on the recommendations of the Working Group, which considered their related importance/share in the total transaction. The selection of the agricultural commodities was done following recommendation of the Sub-group constituted for selection of agricultural commodities. In case of tea, coffee, tobacco etc., the views of respective Boards, namely, Tea Board, Coffee Board were considered. Selection of the items in the mineral sector has been made along the lines suggested by Indian Bureau of Mines.

Demographic Statistics System

Main sources of Demographic Data in India are:

i) Population Census
ii) Civil Registration System
iii) Sample Surveys
iv) National Sample Surveys
v) Sample Registration System
vi) Health and Family Welfare Service Statistics
vii) Other Ad-hoc Surveys

i) Population Census

Population Census is the most comprehensive source of basic demographic characteristics. It provides data on size, structure and growth of population, etc even for small administrative areas. However, one of the most important drawbacks of census as a source of vital statistics is that in India the Census is only decennial operation. Hence the data would be available once in ten years only and this is not sufficient for vital statistics that may respond to policy and program initiatives much faster and would require closer monitoring. The first census was conducted in India in 1872 and the country has an unbroken series of census every ten years from 1881 and the latest census was conducted in 2001. The various data items collected in different census can be categorised into,

- Demographic and social characteristics;
- Educational characteristics;
- Migration characteristics;
- Economic characteristics; and
- Other characteristics.

Demographic and social characteristics

The important items covered under demographic and social characteristics are sex, age, marital status, religion, caste/tribe, literacy, mother tongue, age at marriage, no. of children born to a married women and children surviving, child births in the previous year to married women.

ii) Civil Registration System

India has a long tradition of registration of vital events for over a century. Registration of vital events was being done even in 1850s under various municipal byelaws. In 1886, the Government of India enacted the Births, Deaths and Marriages Registrations Act to provide for voluntary registration of such events throughout the country. Though the Britishers and other foreigners took advantage of the Act, because of the voluntary nature of the provisions of the Act, it had very little application and remained a dead letter. Apart from the above act there were various acts, rules and bye-laws under which registration was being done in various parts of the country. In some areas of the country registration was compulsory while in many other parts it was voluntary. For example, The Bengal Registration of Births and Deaths Act, 1872 provided for compulsory registration in the then Bengal Province. Many states had enacted similar laws even after independence providing for registration of vital events, though the procedures were not uniform. On the basis of recommendations of various committees and with the consent of State governments a central law "The Registration of

Births and Deaths Act" (RBD Act), for compulsory registration of births and deaths in the country was enacted in 1969.

RBD Act enacted by the Parliament is implemented at the field level by the State Government provides statutory authority at the center and in each State. The Registrar General, India coordinates and unifies registration activities across the country. The Chief Registrar of Births and Deaths appointed by the State Government is the executive authority for registration within each state. The Act Provides for appointment of District Registrars and registrars for each local area like village, panchayat, etc.

Under the Act births and deaths are to be registered at the place of occurrence. In case of births/deaths occurring in a household, the head of the household and in case of those occurring in hospitals/ medical institutions, the medical officer – in charge are responsible to report the event to the registrar. All events are to be reported within 21 days after which payment of fine and certain formalities will have to be completed for registration.

The following data items are collected during registration

Births/still births:

Characteristics of the events: date of registration, date of birth, sex, place of birth, order of live birth, type of attention at delivery, birth weight

Characteristics of parents: literacy, occupation, religion, age of mother at birth and marriage.

Deaths:

Characteristics of the events: date of registration, date of death, place of death, cause of death, type of medical attention received.

Characteristics of the deceased: age, sex, marital status, religion, nationality, occupation.

Evaluation :

Despite having statutory provisions, registration data are deficient due to incomplete coverage and under-registration. Extent of incomplete coverage and under-registration varies from State to State. Only about 60 per cent of the births and 51 per cent of the deaths are being registered a the national level. However, birth registration is almost complete in the major states of Kerala, Punjab and West Bengal apart from the State of Goa and union territories of Lakshadweep and Pondicherry. On the other hand death registration is incomplete in most places.

National Sample Survey(NSS)

Information of demographic characteristics of the population on nation-wide scale was collected for the first time in NSS in 1951 through a single round retrospective survey. During 1958-68 population survey become an integral part of the NSS annual rounds; but were discontinued thereafter. Survey on population, family planning and morbidity was conducted in 28th round of NSS (October 1973- June 1974). A survey on births, deaths and population was conducted in 39th round (January 1984- June 1984). The last survey was conducted in 52nd round of NSS during (1995-96) which collected the information on Maternal and Child Health.

Evaluation:

One of the inherent limitations of a retrospective sample survey is the inability of the respondent to recall all events which occurred in the household during the reference period. Though there are methods for adjustment of recall lapse, there does not appear to be a

consistent pattern of errors over space and time; hence universally valid estimates of births and deaths are not available.

iii) Sample Registration System (SRS)

SRS being conducted by the Office of the Registrar General, India since 1964-65 on a pilot basis and became operational on full scale from 1969-70. It is one of the largest demographic sample household survey covering 1.2 million households and about 6.05 million population on a continuous basis. The main objective of SRS is to provide reliable annual estimates of birth and death rates at the National and State levels. SRS also provides estimates of Infant Mortality Rate (IMR) and Total Fertility Rate (TFR).

SRS uses a dual reporting system for collection of data. A resident enumerator in the selected sample areas continuously collects the data on vital events and sends a monthly report. An independent half yearly survey in the same units is also done and the events enumerated are matched. The unmatched events are field verified for arriving at an unduplicated count of the events.

Data on Causes of Death through SRS

Data on causes of death are also collected through SRS since 1999. Earlier the data on cause of death were collected through selected Primary Health Centres in rural areas only. The objective of the survey was to build up statistics on "most Probable Causes of Death" for rural India, States and Union territories, using "Lay Diagnosis Reporting (Post Death Verbal Autopsy)" method through post death enquiry.

Evaluation

A systematic and coordinated effort was made to measure the extent of under-enumeration of events through intensive field enquiry by senior level personnel.

- The under-estimation during 1980-81 was of the order of 3 percent, both for birth and death rates.
- In 1985 the under-estimation was 1.8 percent in birth rate and 2.5 percent in death rate.

iv) Family Welfare Statistics

The main sources of statistics on family planning and maternal and child health (MCH) services are:

- The administrative statistics released by Evaluation and Intelligence (E&I) division of Department of Family Welfare,
- Ad hoc surveys on family planning conducted by various research institutions,
- Base-line surveys in the area projects.

Family Welfare & Health Statistics

Evaluation and Intelligence Division in the Ministry of Health and Family Welfare is responsible for collation of data collected in respect of family planning programmes. The flow of information is from PHC to district family welfare bureau to State bureau and finally to E&I Division at national level. At the grass root level, each family welfare centre is maintains a register of eligible couples. The register contains information on name, age, no. of living children, age of the youngest child, pregnancy status, use of family planning methods and results of follow-up etc. The data collected are published annually.

v) Ad-hoc surveys:

Population research centres, established to undertake research in family planning communication and population studies also undertake surveys in specific areas. Knowledge, Attitude and Practice (KAP) surveys on Family Planning were conducted by Operation Research Group on contraceptive prevalence in 1974 and 1980.

Base-line surveys:

To give a fillip to the national family welfare programme particularly in the backward areas of the country, area projects were introduced the 15 states covering 67 districts of the country. The base-line surveys were contemplated as bench mark surveys to provide feedback on various components. These surveys provide useful data on various parameters of family planning programmes, infrastructure facilities, level of knowledge, attitude and practices, utilisation of health facilities etc.

Health Statistics:

At the national level, the Central Bureau of Health Intelligence (CBHI) and the Directorate General Health Services (DGHS) collects, compiles and disseminates health statistics. The data for some of the diseases, like malaria, cancer etc., are collected from the concerned programme offices. CBHI brings out an annual publication, *Health Statistics of India*. The publication presents data on institutional facilities, morbidity and mortality statistics, data on preventive measures for communicable diseases and on related national programmes.

CBHI attempted to establish an integrated management information and evaluation system (MIES) for the entire Health and Family Welfare programme in the country. Though more than ten years have passed but the receipt of various reports under the programme and the quality of information are still far from satisfactory.

National Family Health Surveys

Two nation-wide Family Health Survey surveys were organised during the last decade to estimate vital rates and the factors affecting them so that detailed analysis of the factors could be done. NFHS-1 was carried out in 1992-93 while NFHS-2 was carried out in 1997-98. Work on the third survey is currently underway. About 3000 women in the reproductive age group were covered in each of the major States in the first two rounds.

Vital rates of different countries*

Countries	Birth rate	Death rate	IMR	TFR	Expectation of life
USA	15	9	7.1	2.1	77
Japan	9	8	3.4	1.3	81
Bangladesh	28	8	66	3.3	59
Philippines	29	6	31	3.5	67
India	26	9	70	3.2	61

* Source: World population data sheet 2001

Population Census

Population census has a long history in several countries dating several centuries back. From giving merely a count of the population of a country as on a reference date, it has grown into a full fledged survey to collect information relating to population on various aspects like social, cultural, economic, demographic, etc. Today most of the countries conduct population census once in a decade and some countries do it once in five years. The United Nations have laid down certain recommendations to be followed in the conduct of population censuses. These relate to periodicity, method of census enumeration, definitions, etc.

Census taking in India started in the 1860s in a non-synchronous manner and the first synchronous census was taken in 1881. Since then there has been a census of the population once every ten years. Over the years the census has grown both in coverage and content with more and more information being collected in every census. As it is the only exercise that can give data on various aspects of the population down to villages and town wards, it has assumed unprecedented importance in the era of decentralized planning. It also combines a housing census being taken at the time of house-listing operations of the population census.

Under the Constitution of India, population census is a Central subject. The Census of India Act, 1948 provides the legal support for taking census in the country. Under the provisions of this Act it is mandatory for the public to provide information asked for in the census. However, the individual information can not be revealed to anyone and can not be used as evidence except for proving that false information was given. It is also compulsory on the part of those deputed for census work to perform such duties as assigned to them. The Act also provides for establishment of a hierarch of officials for conducting the census.

Various stages of the census:

The conduct of population census involves various steps, the most important being,

a) Planning
b) House-listing
c) Enumeration
d) Data Processing
e) Data dissemination.

The planning stage would include decision on the topics on which data are to be collected in consultation with the data users, finalising concepts and definitions of various data items, designing the schedules and their pre-testing, preparation of instructions for enumerators, identifying the enumerators and their training, preparing a frame of the enumeration areas, printing of the schedules, etc. Considering the continuous changes in the administrative structure in the country in terms of districts, tahsils, towns, villages, etc., one of the important steps is to freeze the administrative boundaries in the country. Preparation of maps of every area is important to ensure that all areas have been covered.

The enumeration of population involves tremendous amount of work considering the detailed questionnaire that is used in census. Hence it is necessary to ensure that the enumerators have comparatively uniform amount of work. The house-listing operations in a census has two main objectives, (a) to prepare a frame of all places where people may be living at the time of census and (b) to assess the workload so that it can be evenly distributed among the enumerators. However, the house-listing operations in the Indian Census has now grown into a housing census where a variety of data are collected about the quality of housing, the amenities available to the household and the assets owned by the household. The data collected during the operations are very useful. The basic information collected

during the house-listing is used to carve out Census Enumeration Blocks (CEB) for the population enumeration.

The population enumeration in India is completely through the interview method. As the educational levels of the population is low in most parts of the country it has not been possible to use the method of sending the questionnaires to the households to be filled in by them and mailed back. Primary school teachers are generally engaged as census enumerators. Another important decision relating to enumeration is whether people are to be enumerated at the place of their residence (*de-jure*) or wherever they are found on the reference date (*de-facto*). In India an extended *de-facto* method is being used. Under this method are covered:

(i) All those who normally stay and are present in that household during the entire period of enumeration.

(ii) Also those who are known to be normally residing and had actually stayed during a part of the enumeration period but are not present at the time of enumerator's visit;

(iii) Also those who are known to be normally residing and are not present at the time of enumerator's visit but are expected to return on or before last day of census; and

(iv) Visitors who are present in household censured by the enumerator but are away from the place(s) of their usual residence during the entire enumeration period. For the purpose of enumeration such visitors will be treated as normal residents of the household where they were actually found during the enumeration period provided they have not been enumerated elsewhere.

The reference date for the Census of India is normally the 1st of March. The Census reference period has been fixed taking into consideration the climatic conditions and the agricultural season so that operations are not affected by adverse weather conditions and there is least mobility among people.

Data processing is the next most important step. Careful planning is required in this case so as to ensure that maximum information can be generated from the data within the shortest possible time at the lowest possible cost. Creation of infrastructure through capital investment has to be optimally planned, as census is conducted only once in ten years. In India a mix of manual and computerized tabulation was being used till 1991 census while in 2001 census, only a few provisional data were generated through manual tabulation.

Data dissemination is another important aspect of population census. Several means of data dissemination are available today, like the printed books containing data, analytical reports, data on computer media, internet, etc. The ultimate objective of the dissemination strategy is to ensure that the data reaches maximum number of users. Simultaneously potential users also need to be educated about the availability of data.

Census of India 2001

The latest census of India was conducted in 2001 with March1, 2001 as the reference date. The preparatory work on census started in late 1997. A data users' conference was conducted in 1998 to assess the requirement of various types of data users including those from government, academic and private sectors. Based on the discussions in the Conference and the internal assessment of the capability of the census to cater to the needs of the data users it was decided to add questions on assets owned by households like television, radio, bicycle, etc. and whether they are availing services like banking in the house-list schedule. Similarly, questions on disability and distance traveled to the place of work were added in the main census schedule.

The concepts and definitions of the questions that were continued from the 1991 census were kept same in most cases so as to ensure comparability. The census schedules however underwent a drastic change in its design. The changes were made to make them usable for adopting latest technology for computerization of data by way of scanning the schedules and using character recognition technology to identify the entries. This necessitated strict standards in printing and reducing the number of schedules. Hence, it was decided to have the individual data captured through a household schedule instead of the two schedules used in the 1971-91 censuses.

During the census operations it was decided by the Parliament to create the three new states of Chhatisgarh, Jharkhand and Uttaranchal (now Uttarakhand). As these states were formed by taking out complete districts from Madhya Pradesh, Bihar and Uttar Pradesh respectively, it did not affect the operations seriously. The country had 593 districts at the time of the 2001 census. The location code structure that were being used in the previous censuses underwent a change in this census when it was decided to adopt permanent location code numbers for the villages that would be kept same in the next census so that comparison of data across census would become easier. A hierarchical code structure was being used earlier whereby the villages were numbered within a tahsil, the tahsil within a district and the district within the state. In the new system all villages in a state are numbered continuously. The location of each village in terms of districts and tahsils are identified by separate directories. This obviates the need for giving new codes to the villages when new districts and tahsils are created before the next census.

Data collection in the 2001 census was done during the conventional method of census enumerator visiting each household to collect information. About 1.8 million enumerators and supervisors were deployed for field exercise. As in earlier censuses, the Census of India 2001 was conduct in two distinct but inter-related phases –one the Houselisting Operations and the other Population Enumeration. The house-listing operations were carried out during April-September 2000 followed by the enumeration during February 9-28, 2001 with a revisional round on March 1-5, 2001. In each state the census operations were looked after by a Director of Census Operations appointed under the Census Act. At the District level the District Collector was appointed as the District Census Officer on an *ex-officio* basis. Tahsildar in the rural areas and executive officer in charge of the administration of the Municipalities/Corporations were appointed as Charge officers. Below them were the census supervisors for each supervisory circle consisting of 5-6 enumeration blocks. The census enumerators generally handled the work of a census enumeration block, except when it was very small. An enumeration block in rural areas is either a full village when the village is small or a part of a village. Normally the population size of an enumeration block in rural areas is kept at about 120 households or 750 people. In urban areas, the enumeration blocks are carved out in such a way that it does not cut across the ward boundaries and has a population size of about 500. The urban areas include the statutory towns and those villages with a minimum population of 5000, a population density of 400 persons per sq. km., and 75% of the workers engaged in non agricultural activities. The areas in the periphery of larger towns and the cities with predominant urban characteristics were considered as outgrowths of the concerned towns/ cities. Contiguous urban areas were grouped together to form urban agglomerations so that data could be made available for the agglomeration as a whole

The house-listing work is conducted with a view to prepare a sound frame for carving out the Enumeration Block and for conduct of population enumeration as well as to generate a comprehensive database on the housing stock across the country, its quality and the living conditions of the occupants. It involves listing and numbering all houses and other structures

along with their use, and the names of the head(s) of household(s) living therein and the number of members in the household. In addition, data on housing characteristics such as material used in its construction, availability of electricity, source of drinking water, type of toilet, assets possessed by the household, etc were also collected. An abridged house-list of each enumeration block containing the house numbers, names of the heads of households, etc. is prepared for the purposes of the main enumeration. This is used by the census enumerator to ensure full coverage. In most states the same person conducts the house-listing and enumeration in a block.

The population enumeration involves visiting every household for collecting information. As this is done over a period of about three weeks there is a likelihood of changes by way of births and deaths. To take care of such changes, a revisional round is conducted immediately after the reference date of March 1. The data collected from each individual includes age, sex, marital status, age at marriage, educational level, attendance in educational institutions, economic activity, migration particulars and information on child births and their survival to married women. The enumerators, apart from collecting information, also prepared summary of the data giving the number of persons, those aged blow 7 years, literates, and workers by gender. These were compiled at the supervisor, charge district and state level to prepare the provisional data at state and national levels that were released from end of March 2001 onwards.

The data processing in 2001 census utilised the latest technology for Intelligent Character Recognition (ICR). For using this technology the forms were specially designed and printed according to specific norms. The use of various languages across the country for data collection made the task more difficult. Instructions to enumerators were modified to ask them to ensure that the numerals are written in the international form of Indian numerals only and that too without any ambiguity so that the computer can recognise them correctly. Only the numerically coded information underwent the ICR process. For all other items Computer Assisted Coding was used. The entire process was decentralised to 15 data Centres spread across the country. The editing and tabulation of data were carried out in Delhi.

The tabulation plan for 2001 census contains about 300 tables showing a quantum jump in the number of tables. This increase is due to the addition of new variables in the questionnaire and new cross classifications. As a household schedule has been used for enumeration and that it has been fully computerized, for the first time the census would be in a position to generate several tables at household level containing information based on the members of the household. The data are being released in the following sets of tables.

a) General Population Tables including Primary Census Abstracts
b) Economic Tables
c) Social and Cultural Tables
d) Migration Tables
e) Fertility Tables
f) Housing Tables
g) Household tables
h) Tables on SC/ST

While the Primary Census Abstracts would contain data for each village and town ward, most of the other tables would be at district/city level. Unlike in the past censuses where sampling was resorted to for preparation of detailed tables, all tabulations except those on NIC and NCO classifications in 2001 census are based on 100 per cent count. The use of ICR technology has made it possible.

Printed books used to be the only mode of data dissemination in census till 1981 census. In 1991 census data were made available on computer media- mostly floppy disks. The availability of cheaper technology and proliferation of computers have resulted in higher demand for data on computer media. Today all census tables are being made available through CDs as soon as they are finalised. Data are also released through internet. Printing of the census tables is being done only for national and state level information. As in the previous censuses analytical publications have also been brought out.

In addition to data collected in the census on the population and houses, information on amenities and infrastructure facilities available at the village level and data on statutory towns have been collected and these would be available through the Village Directory and Town Directory respectively.

Civil Registration System in India

Introduction

"Civil Registration System" is a continuous and permanent recording of occurrence of vital events along with their particular characteristics as per the legal provision of the country made for this purpose. In India, the Registration of Births and Deaths Act, 1969 provides for registration of births and deaths.

In India, the system of civil registration has been in operation for more than a century. With the enactment of the central legislation known as Registration of Births & Deaths Act, 1969, reporting and registration of births and deaths has been made compulsory throughout the country. There are more than 2 lakhs registration centres in rural areas and around 4000 in urban areas.

A good Civil Registration System (CRS) can generate reliable data on births and deaths. Apart from providing vital rates, an effective and reliable CRS has to form the basis for other public utility services, which have a bearing on date of birth or death. Being creating a time series, it can be effectively used for better targeting of Health, Family Planning and Child and Maternal Care Programmes. In the context of decentralised planning and demand of statistics at local area level, civil registration data have become of special importance as sample surveys have a limitation in providing estimates for smaller areas. Therefore, registration statistics on births/deaths are the basic tools at local area level for monitoring population growth and effects of health measures.

Civil Registration System is the only source which can provide birth rate, death rate, stillbirth rate, infant mortality rate and maternal mortality rate at all administrative level.

Though the registration of births and deaths has been made compulsory under the Registration of Births & Deaths Act, 1969, yet there is substantial under-reporting of the vital events, mainly on account of lack of awareness among the general public on the need an importance of registration of births and deaths and lack of procedural knowledge on the part of the registration functionaries especially at the lower level reporting, low levels of literacy, lack of proper communication etc. It is estimated that for the country as a whole, only 64% of the births and about 58% of the deaths are getting registered.

Registration Hierarchy

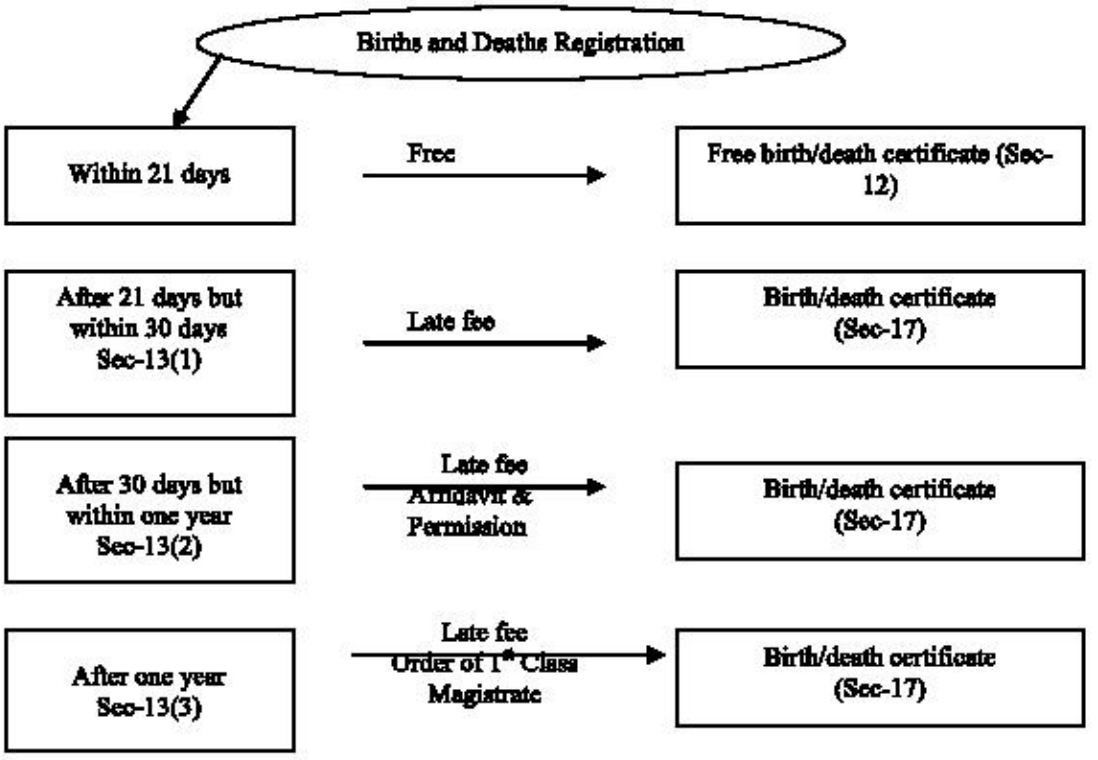
The RBD Act, 1969 provides for the appointment of certain functionaries at various levels:

(i) Registrar General, India at the national level;

(ii) Chief Registrar, Additional Chief Registrar/Deputy Chief Registrar at the state level;
(iii) District Registrar and Additional District Registrar at the district level;
(iv) Registrar for a local area, such as the area under the jurisdiction of a panchayat, municipality, municipal corporation or other local bodies;
(v) Sub-Registrar for specified areas within the area under the jurisdiction of a Registrar.

For a better coordination and implementation of the CRS related activities, the Director/Joint Director/Deputy Director of Census Operation in each state has been designated as Joint Registrar General/Assistant Registrar General.

3. Registration Procedure



Statistical Reporting System

Flow of Returns from the registration centers:- The Registrar after completing the process of registration is required to prepare and submit the statistical returns related to events registered during the previous month along with the Monthly Summary Report in form no. 11, 12 and 13 meant for live births, deaths and still births respectively and send them to the District Registrar's Office (or any other office assigned for this purpose) for statistical processing on or before the 5th of the following month.

Flow of Returns from the District Registrar Office:- District Registrar Office is supposed to scrutinize and compile all the registration information received from all the registration centers under his jurisdiction. A consolidated statistical report is required to be submitted to the office of Chief Registrar of the State on monthly basis on or before the 10th of the month.

Flow of Returns from the State Headquarters:- Chief Registrar is required to scrutinize and compile registration data received from all the districts and prepare a statistical report containing data for each calendar year. The Chief Registrar is also required to prepare an annual report on the Working of the Act. These two reports are to be submitted to their State Government by the 31st July of the following year.

Chief Registrars of the states are also required to send to the Registrar General, India
(i) Monthly returns within two months after the expiry of the month to which the return relates.
(ii) The Annual Reports within nine months after the end of the calendar year.

Annual Reports at the National Level: Based on the annual reports prepared by the states, the Office of the Registrar General, India is bringing out annually a comprehensive Annual Report on Working of the RBD Act, 1969 and also a statistical report entitled 'Vital Statistics of India based on the Civil Registration Data'. The Annual Statistical Report prepared by the ORGI is a series of reports that have been brought out starting with 1958. The RGI is required to submit to the Central Government the annual report on the Working of this Act.

CRS Tabulation Plan under Revamped System of Registration of Births and Deaths

In order to reduce the paper work and to speed up the process of registration and the tabulation of CRS data, the registration system has been revamped. The procedures for registration of births and deaths have been modified to some extent and the forms redesigned. In order to cater the need of planners, demographers, researchers etc., most of the tables in the tabulation plan of registration of births and deaths data under revamped system are prepared based on the information of place of usual residence of mother in case of births and place of usual residence of deaths in case of deaths.

The details of the forms under the revamped system of registration of births and deaths are as follows:

Form No. 1: Birth Report Form;
Form No. 2: Death Report Form;
Form No. 3: Still Birth Report Form;
Form No. 4: For Medical Certification of Cause of Death (for Hospital Patients);
Form No. 4A: For Medical Certification of Cause of Death (for Non-Institutional Deaths);
Form No. 5: Birth Certificate;
Form No. 6: Death Certificate;
Form No. 7: Birth Register;
Form No. 8: Death Register;
Form No. 9: Still Birth Register;
Form No. 10: Non–Availability Certificate;
Form No. 11: Summary Monthly Report of Births
Form No. 12: Summary Monthly Report of Deaths
Form No. 13: Summary Monthly Report of Still Births

The following tables are prescribed in the tabulation plan for CRS under the revamped system of registration of births and deaths:

a. Two tables (2), Table A-1 & Table A-2 for monitoring the returns of receipt from registration units separately for Rural and Urban Areas respectively.

b. Twenty three (23) tables for births (from Table B-1 to Table B-23), twenty one (21) tables for deaths (from Table D-1 to Table D-21) and seven (7) tables for stillbirths (from Table S-1 to Table S-7) are prescribed in tabulation plan.

c. Out of 23 tables prescribed for births in the tabulation plan, only two tables, namely, Table B-1 and Table B–3 will be prepared by taking all registered births in an area during a calendar year. Other tables of births will be prepared only for those births whose the mother is a resident in the state and birth occurred in the year for which the report has to be generated.

d. Out of 21 tables prescribed for deaths in the tabulation plan, only two tables, namely, Table D-1 and Table D–3 will be prepared by taking all registered deaths in an area during a calendar year. Other tables of deaths will be prepared only for deaths of deceased whose place of resident in the State and the death occurred in the year for which the report has to be generated.

e. Out of 7 tables prescribed for births in the tabulation plan, only one table, namely, Table S-1 will be prepared by taking all registered still births in an area during a calendar year. Other tables of still births will be prepared by taking the place of usual residence of mother into consideration and the year for which the report has to be generated.

Determination of performance of the States in implementation of the RBD Act, 1969:

The performance of a State/Union territory in implementation of the RBD Act is generally, judged on the basis of two important indicators:

(i) Level of reporting from the registration centers
(ii) Level of registration (LOR)

Level of reporting from the registration centres:

It indicates the efficiency of the registration machinery. The percentage of returns receipt out of total expected returns from registration centres is used as an indicator of the level of reporting from the registration centres. It is calculated separately for births, deaths and stillbirths. Mathematically,

Level of reporting for birth summary report
= (Number of summary monthly reports of births received from registration centres) / (Number of registration centres * 12) * 100

Level of reporting for death summary report
= (Number of summary monthly reports of deaths received from registration centres) / (Number of registration centres * 12) * 100

Level of reporting for stillbirth summary report

= (Number of summary monthly reports of stillbirths received from registration centres) / (Number of registration centres * 12) * 100

Level of registration (LOR):

It actually determines the performance level of a State/ Union territory. The percentage of the registered live birth/death to the estimated total number of events using SRS rates are used as an indicator of the level of registration as no other reliable estimates of number of live births and deaths are available. Mathematically,

LOR of Births = (Number of registered births) / Number of estimated births based on SRS rates) * 100

LOR of Deaths = (Number of registered deaths) / Number of estimated deaths based on SRS rates) * 100

The level of registration so calculated is very high in smaller States/Union territories either due to the lower estimates of vital events than the actual and / or due to the estimated events based on SRS related to usual residents only while events are to be registered at the place of occurrence irrespective of the place of residence.

Format of the Report on the Working of the Act

Every Chief Registrar of Births and Deaths has to submit to the State Government a report on the working of the Act in the state every year by July 31 of the following year along with a statistical report as prescribed in the rules.

The format of the Report on the Working of the Act is as under:

1. Brief description of the State, its boundaries and revenue districts.
2. Changes in Administrative Areas.
3. Explanation about the difference in Areas.
4. Changes in Registration Areas-Extension.
5. Administrative set up of the registration machinery at various levels.
6. General response of the public towards this Act.
7. Notification of births and deaths.
8. Progress in the medical certification of cause of death.
9. Maintenance of Records.
10. Search of births and deaths register for issue of certificates.
11. Delayed registrations.
12. Prosecutions and compounding of offences.
13. Difficulties encountered in implementation of the Act.
    i) Administrative
    ii) Others
14. Orders and Instructions issued under the Act.
15. General remarks.

However, it has been observed that many of the State reports are very brief and sketchy and contain very little factual information about the status of registration and steps undertaken by the State Governments and Chief Registrars for improvement of the situation. They mainly describe the procedural aspects instead of giving information on the steps taken for improving the system.

Few states are not preparing and submitting the Working of the Act Report and the Statistical Report as prescribed under the sections 4(4) and 19(2) respectively of the RBD Act, 1969 separately. These two reports should be prepared separately as the Report on the Working of the Act is mainly for official use for taking steps streamlining the registration system whereas, the Statistical Report is for public and is required to be published as per the provisions of the Act.

Few observations on working of the system in the State

Some of the main observations made by the officials from the office of the RGI during field inspections are:

a) Non-display of signboard and timings regarding Registration of Births and Deaths in most of the places.
b) Faulty design of the reporting forms
c) Use of wrong forms/ wrong printing of the reporting forms
d) Training to registration functionaries about the various provisions of the RBD Act 1969 and the State Rules made there under and on basic demographic concepts is required
e) Incorrect procedures in reporting the events as well as incomplete filling up of the reporting forms were observed.
f) Registration at wrong place
g) Non-registration of events on the basis of the reports directly received from the medical institutions.
h) Inconsistency in the cause of death reported in the death report form from that reported in the Medical Certification of Cause of Death.
i) Registrars do not sign the registers
j) Corrections in the Register without following the correct procedure
k) Proper maintenance of the current registers is required
l) Preservation and safe keeping of old records are required
m) Non-submission of monthly reports in time by the registrars and even in many cases no efforts are being made by the district/state authorities to collect them.
n) Non-issuance of birth/death certificate (free of charge) under Section 12 of the RBD Act, 1969 in most of the states.
o) Errors/Non uniform procedure observed in giving registration number
p) Lack of communication to Registrar

Sample Registration System (SRS)

Background

The Office of the Registrar General, India, initiated the scheme on sample registration of births and deaths in India popularly known as Sample Registration System (SRS) in 1964-65 on a pilot basis and on full scale from 1969-70 mainly due to lacunas in the civil registration system. SRS is the largest continuous demographic sample household survey in the world covering about 1.3 million households and about 7 million population.

Survey format

SRS is a dual record system with two independent field functionaries. A resident Part Time Enumerator (PTE) continuously enumerates the events in the selected sample units. An independent six monthly retrospective survey is done by a full time supervisor from the Directorate of Census Operations. The events collected by both are matched and unmatched events are verified in the field. This procedure helps in getting a better count of the number of events without the drawbacks of the single round retrospective surveys

Objectives and scope

The SRS was initiated with the objective of providing reliable estimates of vital rates in the absence of such estimates from any other source including the Civil Registration System. Over the years the scope of the survey expanded and now it provides reliable estimates of

- Birth and death rates
- Infant mortality rate
- Various other measures of fertility and mortality including child mortality, total fertility, etc.

The data collected in SRS is also used to estimate other parameters like expectation of life periodically.

Level of presentation of data

The data on vital rates -birth and death rates- are presented for all states and union territories separately for rural and urban areas. Estimates for IMR are presented annually for major states while for other states and union territories, it is presented as moving averages for three years so that fluctuations due to small sample can be taken care of. In future, it will also provide reliable estimates of BR, DR & IMR at NSS Natural Division level for rural areas.

Organization of work

SRS is implemented by Directorates of Census Operations (DCOs) except Kerala and Maharashtra where rural units are handled by Directorate of Economics & Statistics. At national level, the Vital Statistics Division of the Office of the Registrar General, India coordinates and supervises the implementation, formulates and prescribes necessary standards, provides necessary instructions and guidance and undertakes tabulation and analysis of data and their dissemination.

Revision of SRS Sample

The sample is changed after every Census so as to take care of changed demographic scenario which may need modifications in the sampling design to make it more efficient. This also helps to give better representation of various sections of the population and to

overcome the difficulties/limitations in the existing scheme. This opportunity is also utilized redesign the survey to meet the additional requirements.

The latest change in the sample has bee made effective from 2004 based on 2001 Census. Simultaneously the Forms and procedures were rationalized with the objectives of,

- Better netting of events
- Making them more user friendly
- For adaptation of latest technology like scanning
- Streamlining the system

Sample design

In Rural areas SRS uses a uni-stage stratified simple random sample without replacement, except for larger villages. If the village selected has a population below 2000 then the entire village is selected. Otherwise the village is segmented into two or more segments and a segment of the village is selected randomly.

In urban areas the census enumeration block is the sampling unit. The urban area in each state is stratified into four strata as below.

Stratum – 1: Towns with population less than 1 lakh
Stratum – 2: Cities with Population 1-5 lakh
Stratum – 3: Cities with Population 5 lakh and above
Stratum – 4: Four metro cities, Delhi, Mumbai, Chennai and Kolkata as separate strata.
(proportional allocation to slum population)

Sample size

The new SRS sample size is designed for estimating IMR at Natural Division level for rural areas as well as urban areas with 10% relative standard error for all major states. In smaller states, the sample size is sufficient to provide estimates of IMR for both urban and rural areas with 15% relative standard error. As births and deaths are more frequent than infant deaths the sample would give estimates of births and death rates with much higher precision. The number of sample units at the national level has been increased from 6,671 to 7,597 with 4,433 in rural and 3,164 in urban areas with effect from 1994.

Unique Identification Code

One of the significant modifications in the field data collection process is the introduction of unique identification code for each individual. This will help in easy storage and retrieval of data, aggregation at different levels, cross-classification of various determinants with fertility and mortality indicators, cohort studies, etc. The identification code comprises two parts, viz., Sample Unit Code and the Individual Identification Code.

Main Components of SRS

Baseline survey of the sample units is conducted after selecting the sample. Continuous (longitudinal) enumeration of vital events is taken up thereafter through the local part-time enumerators appointed for this purpose. Independent half yearly survey for recording births and deaths that occurred during the previous six months is taken up thereafter. The events reported by both the sources are then matched in the office and field verification of unmatched and partially matched events is taken up.

Estimation Procedure

The estimates of population, live births, deaths and infant deaths are obtained using unbiased method of estimation. Estimates of birth, death and infant mortality rates as well as other fertility rates are obtained as the ratio of the corresponding estimates.

Survey of causes of death through SRS

The ORGI initiated a scheme, viz., "Model Registration Scheme" (MRS) during the 1960's for collection of data on causes of death covering only the rural areas. This Scheme was renamed in 1982 as "Survey of Causes of Death (Rural)" (SCD). Implementing Agencies of SCD (Rural) were State Directorates of Health Services. On account of restricted coverage of SCD to only rural areas and other operational problems, the scheme was integrated with SRS from 1st January, 1999. In SRS the data on causes of death is being collected from both rural and urban areas. The causes of death data are collected on 79 specific cause groups, which are classified into 19 major cause groups as per International Classification of Diseases (ICD) – 10th Revision.

New Initiatives in SRS on data on causes of death

The forms for collecting data on symptoms were revised based on the existing experience of WHO, Chinese Surveillance System and other international and national studies. Now there are forms providing a structured set of questions and also having provision for narrative descriptions. There are separate forms for Neo-Natal deaths, Childhood deaths, Adult deaths and Maternal deaths.

Partnership has been established with premier Institutes like CGHR (University of Toronto), NIMHANS (Bangalore), PGI (Chandigarh), ICMR, TIFR (Mumbai), Medical colleges of India, ERC (Chennai), for

- Training/Refresher Training to RGI Staff on verbal Autopsy
- Conducting VA in 10 percent resample units
- Assignment of causes of deaths (double assignment)
- Quality Control
- Epidemiological analyses

The cause of death is assigned by a health professional. About 10% of the cases are independently assessed by different person so as t improve the quality of the data.

Publications from SRS

1. SRS Bulletin (six monthly)
2. SRS Statistical Report (Annual)
3. SRS Based Abridged Life Tables (Biennial)
4. Occasional Papers such as 'Below State Level Estimates, 1995-1999'
5. Report on Causes of Deaths

- What is Censes?
- Concepts of Design of Experiments
- All Sampling Methods
- Some popular Designs e.g. CRD, RBD and LSD
- Analysis of Variance ANOVA
- One way and Two way ANOVA
- Missing Value ANOVA
- Vital Statistics,
- Demography
- Mortality and Its Measures
- Fertility and Its Measures
- Migration and its measures
- Infant Mortality Rate
- Immigration
- Growth Rates
- GRR and NRR
- Index Numbers
- Fishers Index Numbers
- Control Charts
- 3 sigma limits
- 6 sigma limits
-