

<b>COURSE CODE:</b>	AEM 505
<b>COURSE TITLE:</b>	Research Methods
<b>NUMBER OF UNITS:</b>	3 Units
<b>COURSE DURATION:</b>	Three hours per week

---

### COURSE DETAILS:

<b>Course Coordinator:</b>	<b>Prof. (Mrs.) C.A. Afolami</b> <i>B.Sc., M.Sc., PhD</i>
<b>Email:</b>	<a href="mailto:carolynafolami@yahoo.co.uk">carolynafolami@yahoo.co.uk</a>
<b>Office Location:</b>	<b>Agric. Econs &amp; Farm Mgt, COLAMRUD</b>

### COURSE CONTENT:

Overview of a Research Proposal, Problem definition and Conceptualization, Hypothesis formulation and Objectives, Analysis of Objectives and formulation of questionnaire, review of related literature, types and sources of data, methods of data collection, conducting surveys, data processing, overview of appropriate methods analysis, result presentation and discussion, referencing, writing a research report, computer application in socio economic research. A research proposal will be submitted as part of the course requirement.

### COURSE REQUIREMENTS:

This is a compulsory course for 500 level students in the university. In view of this, students are expected to participate in all the course activities and have minimum of 75% attendance to be able to write the final examination.

### READING LIST:

1. Eboh, E.C. (2009). *Social and Economic Research: Principles and Methods*, Africa Institute of applied Economics, Enugu Nigeria.  
Website: [www.aiaenigeria.org](http://www.aiaenigeria.org), First published 1998.
2. Ferber, R and Verdoorn, P.J. (1980). *Research methods in Economics and Business*. The Macmillan Company, Canada. 573p
3. Kothari, C.R.(1978). *Quantitative Techniques*, Vikas publishing house PVT Ltd. Pp.64-154.
4. Yamani,Taro (1973) *Statistics: An Introductory Analysis*. Harper International Edition, Time printing, 1130p

### LECTURE NOTES

#### INTRODUCTION

- Scientific writing are research papers, text books, journal articles, monographs, technical papers, their features are such that they are:
- Not developed in isolation from preexisting works

- Knowledge is dynamic & interactive
- No book is an island
- All works must be placed in context of existing knowledge
- Every work requires references to the writing & publications of other authors
- It is necessary to identify the works. This is done by making reference to them both in text and in a list at the end of assignment
- The practice of acknowledging authors/sources of ideas & information used in work/essay/report/assignment is called referencing.
- It is the bibliographic listing of information sources at the end of any academic work
- Information sources are books, articles, internet/print & electronic

### **PURPOSE**

- Reference can be a paraphrase or exact quotation
- Using numerous references demonstrates wide reading to identify & use current thinking, use ideas expressed by others to reinforce own argument
- Shows the breadth/depth of your research
- Shows readers source of your information
- Strengthens your academic argument
- Allows readers to consult your sources of information
- To avoid plagiarism

### **REFERENCING STYLE TYPES**

- 1.American psychological association (APA)
- 2.Modern language association (MLA)
- 3.Havard
- 4.Australian journal of physiotherapy (AJP)
- 5.Foot notes
- 6.Turabian
- The first four are author-date system while the last two are numerical-system

### **RULES OF APA REFERENCE STYLE**

- For 3 or less number of authors citation, spell out all authors on all occurrences
- For more than 3 authors, name all authors on first mention, and then subsequently use "*et al*"
- Use commas to set off a reference in parenthical comment e.g. (Partrick, 1993)
- Join names in multiple-author citation with and or ampersand (&) e.g Smith and Sarah (1990) or smith & Sarah (1990)
- If a group is readily identified by its initials, spell out only at first mention e.g (.cocoa research institute of Nigeria [CRIN], 1990). Then subsequently use-(CRIN, 1990)
- If author is unknown or not specified, use the first few words of title, e.g. ("study finds," 1992)
- For old works, cite translation or original and modern copyright dates if both are known e.g (Aristotle, trans,1931) or (James, 1890/1983)
- Always give page numbers for quotations, e.g (cheek & buss, 1981, p.332) or (Shultz, 1989, chap.3, p.5)

### **EXAMPLES**

- Samo and Porter (1997) point out that 'language involves attaching meaning to symbols" (p.160) or Attaching meaning to symbols is considered to be the origin of written language (Samo & Porter, 1997,p 160)
- Peter (as cited in Samo & Porter, 1997) states that language must be first understood as symbols.
- American Psychiatric Association [APA], 1990) next citation will be (APA, 1990)
- For authors more than 3: at first mention in text, state as- (Adams, Baca, Smith and Jones, 2000) subsequently state as (Adams et al., 2000)

### **IN BIBLIOGRAPHY**

- At the end of research paper, you are required to provide the full reference for each source.

- Reference must be listed in alphabetical order
- Each reference should include 4 elements; author/editor, date, title and publication information- Publisher and location/place

### EXAMPLES

- Samo, L.A. and Porte, R.E. (1997) *Intellectual Communication: A Reader* (8th ed.) Belmont, C.A: Warsworth.
- American Psychiatric Association.(1990). *Diagnostic and Statistical Manual of Mental Disorders*(3rd ed.) Washington, D.C.; Author
- "Author" is used above when author and publisher are identical
- In a situation where the book has only editors, write their names and add (Eds)-eg : Samo, L.A. and Porter, R.E.(Eds)(1997).....

### MODERN LANGUAGE ASSOCIATION REFERENCE STYLE

- Requires all titles to be in italics or underlined
- In the body of work, MLA citation requires only enough information to find the source, usually author and date in parenthetical reference.
- You only mention authors' name, no initials-in text
- The list of sources consulted appears at the end with entries in alphabetical order
- If there are no authors, use work title
- Example-bibliography listing
- Berkman, Robert I. *Find Fast: How to uncover expert information on any subject*, New York: Harper Perennial, 1994
- 2. For no author or editor work, the title is used as element of citation. Do not use anonymous or anon, rather alphabetize entry by first main word of title, *The Dictionary of Business*..

### FOR WITHIN TEXT REFERENCE

- Afolami (2008) found that productivity can be improved by efficient use of resources...OR
- Resource use efficiency study would facilitate improved productivity (Afolami, 2008).

### HAVARD REFERENCE STYLE

- it is an author-date system of referencing
- Its used within text and in the list of bibliography
- The reference list is arranged alphabetically by authors family name, followed by comma, then initials (no bracket), title of book in italics, followed by comma, publisher and place of publication, with capital letter for first words only.

#### • EXAMPLE:-Book

- Daly, L. Mohay, H. and Edward, H. 2003, *Mother's involvement in Caring for Premature Infant*, Elsevier, Sydney.
- Example for journal article:
- Davis, L, Mohay, H and Edward H 2003, 'Mothers' Involvement in Caring for Premature Infants' *Journal of Advanced Nursing*, Vol.2, No.6, pp34-78
- *Et al* is used in the text for authors more than three
- All authors are listed in reference list in alphabetical order

### AUSTRALIAN JOURNAL OF PHYSIOTHERAPY REF STYLE

- Author-date style of referencing in case of text paraphrase, quote or copy
- There are 2 components of ajp style which include author and date (sometimes page number) in brackets
- All detail of reference list must be at the end of the assignment
- In-text referencing in ajp ref style;-when you cite references regardless of whether you paraphrase, summarize, quote or copy, you should include author's surname or name of editor or organization responsible, year of publication and page numbers where available.

### EXAMPLE OF AJP REF STYLE

- A recent study (Lim and King 1998) found.... OR Lim and King (1998) found...
- For 3 or more authors-
- (Boyd et al 2004) have indicated that....
- In reference list, the reference is as follows;
- Patton M.O and Krust R.A (1987): Qualitative Evaluation and Research Methods, Newbury Park, CA: Sage Publication
- For multiple authors-
- Adams M.A, Broch N, Burton K, and Dolan P. (2002): The biomechanics of Back Pain, Edinburg: Churchill Livingstone

### FOOT NOTES

- IT IS A NUMERICAL SYSTEM OF REFERENCING
- YOU USE CONSECUTIVE NUMBERS CALLED IDENTIFIERS WITHIN TEXT STARTING FROM 1 AND SO ON, ie 1, 2, 3...
- Numbers are used to identify information that have been referenced
- Each note identifier is listed at the bottom or foot of each page with relevant bibliography detail
- Footnotes are usually in smaller font
- Identifiers are placed at the end of quotes
- They can be written as superscripted arabic numbers at foot or in-text footnotes

### EXAMPLES

- The conditions necessary for germination was given by (1, 1987)
- -----
- 1 .Adebola A.O (1980).The Essentials of Biology.Acada Press, Ibadan
- The conditions necessary for germination was given1
- -----
- 1 .Adebola A.O (1980).The Essentials of Biology.Acada Press, Ibadan
- 

### COMMON MISTAKES IN REFERENCING

- Mix-up of different styles
- Mis- use of principles of referencing- put quote in inverted commas and quote verbatim
- Punctuation and capitalisation (some style punctuate others do not)
- Maintain a particular style
- Authors should provide required information.
- Golden rule is consistency, i.e use one style and maintain consistency throughout work

#### 1. OVERVIEW OF RESEARCH PROCESS: WHAT IS RESEARCH?

Investigation or experimentation aimed at the discovery and interpretations of facts, revision of accepted theories or laws in the light of new facts or practical application of such new laws. Definition implies that research can generate different outcomes-outcome could be revalidating existing knowledge, generation of new knowledge, adaptation of existing knowledge in new domains and revision/improvement of existing knowledge based on newly revealed facts research

- Manipulation of things, concepts or symbols for the purpose of generalising, to extend, correct or verify knowledge-whether that knowledge aids in construction of theory or in practice of an art.
- An inquiry into the nature of, reasons for, and the consequences of any particular set of circumstances, whether these circumstances are experimentally controlled or recorded as they occur-(kothari, 1990)
- The process of systematic inquiry by which humankind increases knowledge of how Things are, why things are the way they are and how they could be improved.
- It is disciplined and verifiable study of a subject aimed at learning new facts and Testing scientific propositions and ideas

- Research is systematized quest for new knowledge. the basic goal is to produce Knowledge and create understanding through scientific and logical reasoning

### **PURPOSE OF RESEARCH**

- To discover answers to questions through the application of scientific procedures
- Main aim of research is to find the truth which is hidden and which has not been discovered yet.

### **ROLE AND FUNCTIONS OF RESEARCH**

- In the past, economic growth was underpinned by traditional factors of production viz land, labour and capital
- Today, the critical drivers of global economic trends are technical knowhow, creativity, intelligence and information.
- Research makes an original contribution to advance existing stock of knowledge by which societies and economies can be better, prosperous and more fulfilled

### **ROLE AND FUNCTIONS OF RESEARCH**

- Research is important to government, business and industry, civil societies and economy as a whole.
- Social and economic progress depends on the quality and utilisation of research evidence by govt policy makers, industrialists, business people and social groups
- Research provides evidence bases for govt social, economic and development policies
- Sound policies are grounded in sound research.
- For govt economic and social policies to be well formulated, they must rely on sound research evidence
- Research contributes information through diagnosis (probing situations and phenomena) and prognosis (predicting future developments)
- for example, research into a community's economic structure, dynamic changes and underlying factors is a significant tool of decision making for policy/program targeting
- Research can evaluate the potential implication of alternative policies and programs.
- Research and policy making are complementary and mutually reinforcing.
- Research can inform policy while learning from policy itself
- Policy research focuses on helping to evaluate alternative policy options and fosters the emergence of superior policy options.
- It is important to note that policy formulation process is a politically competitive process; policy makers may choose options that are not superior alternative if not properly guided.
- Facts and figures from research are also useful to civil society in carrying out advocacy, civic education and public enlightenment
- Research is also important in solving operational and planning problems in business and industry
- Business decisions benefit from operations research, market research and motivational research.
- Operations research involves the use of mathematical, logical and analytical techniques in solving business objective of optimisation-cost minimisation and profit maximization
- Motivational research involves human behaviour towards consumption and products.
- Market research contributes to identifying policies for purchasing, production and sales.
- What research does is to give business people, entrepreneurs and industrialists a logical and systematic basis for business decisions, rather than relying on guesswork.
- Research also helps to satisfy intellectual yearnings of man for better understanding of social phenomena and economic relationships and processes
- In social science, one dimension of research pursues knowledge for its sake
- As a science, research has the responsibility to accumulate knowledge for more accurate prediction of human behaviour
- On another dimension, research can produce knowledge that contributes to resolving practical real-life problem of human relations.
- research can be broadly classified as incidental –as in auditors investigation/inquiry panel into an issue or systematic as in day-to day activities, i.e. Continuous stages of reseach process
- In research, what differs is the object of study.

- Science is united in philosophical foundations, objectives, logic and methodology
- For example, physics is concerned with matter and energy. Biology deals with life processes of plants and animal. Sociology is concerned with social behaviour and relationships. economics is concerned with resource allocation and choice problems of man.
- Despite disciplinary differences between fields, of scientific study, they share same research process

### **Research process**

- It involves a sequence of mutually-inclusive and reinforcing steps and procedures for Acquiring reliable knowledge
- There are many diverse ways to describe research process but there are common and essential elements and stages of the process.
- Generally as a process, research is a sequence of activities or steps leading to an end Product.
- It has a procedural and step-wise character one view point of research process:

1. Formulating the research problem
2. Extensive literature survey
3. Developing hypotheses
4. Preparing the research design
5. Determining the sample design
6. Collecting the data
- 7 Execution of the project
8. Analysis of data
9. Hypotheses testing
10. Generalisation and interpretation, and
11. Preparation of report and presentation of the results

### **Another description of research process (abridged)**

1. Research design
2. Sampling
3. Data collection
4. Data analysis
5. Report

### **DISCUSSION OF CONSTITUENT STEPS**

- FORMULATING THE RESEARCH PROBLEM
- It means stating the general research topic as an analytical object
- Research problem statement is a distillation of the research subject into feasible investigative questions that are precise, concrete and unambiguous.
- The research problem can either relate to the state of nature or to relationships between variables.

### **EXTENSIVE LITERATURE SURVEY**

- Involves reading and studying materials related to the research subject or problem.
- Materials can be found in journals, books, magazines, periodicals, monographs, seminar papers, conference proceedings

### **RESEARCH PROCESS STAGES**

- DEVELOPMENT OF WORKING HYPOTHESES
- This is a tentative assumption made for the purpose of testing its logical validity .
- The role of hypothesis is to guide the researcher to be focused on specific area of inquiry and keeping him on track.
- The hypothesis determines the data to be collected and the type of analysis to be done.

- **PREPARING THE RESEARCH DESIGN**
- It involves decisions about the units of study, methods of study and techniques to solve the research problem.
- The research design is the conceptual structure upon which the research would be carried out.
- Research designs can be broadly categorized into experimental and non-experimental

### **RESEARCH PROCESS STAGES**

- **DETERMINING SAMPLE DESIGN**
- Researcher has to decide the way and means of choosing the units of the study-individuals, households, enterprises, communities, societies etc
- The sample design is the plan for selecting the representative group from the population of study.
- **COLLECTING DATA**
- Involves observation, experimentation and field exercises.
- There are different types of data just as there are different methods and techniques of data collection
- Choice of type of data and data collection method depends on a variety of factors, the research problem, working hypothesis, time and money available for the research
- **ANALYSIS OF DATA**
- Data analysis requires several related operations involving data organization, data exploration, data tabulation and statistical testing of hypothesis.
- **HYPOTHESIS TESTING**
- Involves the researcher to answer the question-do the facts support the hypothesis or not?
- Where there are no hypotheses initially, the generalizations emanating from research could form the basis of hypothesis in future research.

### **• GENERALISATIONS AND INTERPRETATIONS**

- Research is significant in the sense that it seeks to arrive at generalizations based on empirical observations and logical validity.
- Generalization could involve explaining the results of research in the context of existing theories or previous studies.
- It is possible for generalization to open new questions that require further research.
- Common feature and attributes that underpin the nature and conduct of research is that research is cyclical and an interactive process.
- Sometimes research is presented in a very simplistic linear succession of steps with definite start and end, this is erroneous
- Research appears more as a spiral or iterative process.
- Stages or steps in research process are not mutually exclusive but mutually inclusive. They are neither separate nor distinct. Usage of a term may vary from place to place and may have several interpretations meaning it may be ambiguous. Often people do not understand what we wish to communicate. We thus need concepts and conceptualization

## **2. CONCEPTS AND CONCEPTUAL FRAMEWORK**

### **DEFINITION**

- A concept is the distinctive meaning of a term.
- A scientist is able to manipulate, study, organize and isolate the properties of objects and give names to such properties by formulating concepts
- A clear formulation of concepts is a basic requirement in the application of scientific method.
- Without a clear definition of concepts or with a vague definition, a researcher ends up not knowing what to look for or unable to distinguish it from other related forms
- A researcher must therefore define the concepts well in advance with rigour that when confronted with reality they either hold or breakdown without ambiguity.

### **CONCEPTUALISATION**

- To conceptualize is to formulate concepts.
- It is to communicate the precise meaning of a term in your own construct of the term.

- Conceptualization refers to the process through which we specify precisely what we mean when we use a term, i.e. the working agreements.
- Suppose we want to find out if men have greater access to employment than women.
- We cannot meaningfully do this without some working agreement on the term-employment.
- Specifying or defining what we mean by a term is the way to conceptualize. Research progresses with the specification

### **END PRODUCT OF CONCEPTUALISATION**

- The end product of conceptualization process is the specification of a set of indicators of what we have in mind, indicating the presence or absence of the concept we are studying
- By conceptualization, the researcher is able to isolate and indicate the properties of objects and give names and identity to such properties.
- We can conceptualize the term employment in several dimensions
- Specifying dimensions and indentifying the various indicators for each of those dimensions are both parts of conceptualization

### **TWO LEVELS OF CONCEPTUAL ORDER**

- 1. NOMINAL and OPERATIONAL
- A nominal definition is the name or label that a researcher assigns to a term, *a priori*.
- It gives distinctive identity to an object or thing for purpose of separating it from the rest.
- To avoid confusion or misinterpretation of a term, the scientist specifies a working definition for the purpose of inquiry.
- An example is to specify socio-economic status of an individual as combination of income and educational attainment, precluding many other aspects e.g wealth/assets/occupational status
- Nominal definition focuses on observational strategy of research

### **CONCEPTUAL OPERATIONAL DEFINITION**

- To identify what is to be observed, the researcher has to operationalise the concept, meaning he should have the working definition of the concept.
- An operational definition is one that specifies exactly how the concept is to be measured
- Example: socio-economic status of a household head
- Income+ Educational status
- Income can be total amount of income in the past one year
- Education status can be possession of primary, secondary or tertiary education/post tertiary education coded 1, 2, 3 and 4

- Conceptual framework embodies the progression from the abstract to the operational
- The formulation of operational definition known as operationalisation, is the specification of variables and indicators
- In research, measurement entails assigning possible values to units of analysis, hence we measure concepts that vary and refer to them as VARIABLES
- We cannot see education but we observe how knowledgeable people are
- We need to specify ways of observing the presence or absence of our concepts

### **SUMMARY OF CONCEPTUALISATION PROCESS**

- Conceptualization process involves 4 steps.
- 1. Conceptualisation
- 2. Nominal definition
- 3. Operational definition
- 4. Measurement in the real world
- 5. Formulating conceptual framework

### **MEASUREMENT IN EMPIRICAL WORK**

- Objects of our study may be an abstract concept or a qualitative property.



- Researchers have to assign numbers or labels to objects of observation whose conceptual properties are the target measurement
- Variables represent the dimensions of a concept
- It is a characteristic of a person, object or phenomenon that can take on different values
- While **concepts** are in the domain of **theory**, a **variable** is **observable and measurable**

### **VARIABLE MEASUREMENT**

- Assignment of values to a variable is called operationalisation
- The requirement is to specify indicators as a measure of the variable.
- No two indicators will measure a given concept/variable in the same way.
- Let us consider measuring income
- Concept: Income
- Variable: Earning from employment
- Indicator1: Amount earned in formal employment in a month
- Indicator2: Amount earned outside formal employment/month
- Indicator3: Indicator 1 + Indicator 2

### **ROLE OF VARIABLE IN MEASUREMENT**

- A variable can be **numerical or categorical**
- **Numerical variables** are those whose values are expressed as **numbers**.e.g.1,2,3,..
- **Categorical variables** are those whose values are expressed in **categories** e.g sex(male,female), occupation, color, status,religion, knowledge.
- These need to be **operationalised** i.e **made measurable**
- You can code them, e.g female=0, male=1

### **TYPES OF VARIABLES IN RESEARCH**

- Variables are specified in research based on their functionality in relation to problem under study viz
- Independent variables/exogeneous/explanatory variables.
- Virtually all studies include background variables of age, sex educational background, socio-economic status, marital status, religion etc so they are sometimes called background variables
- Dependent variables/endogeneous variables are those used which are used to measure or describe the problem or phenomenon. They are caused by other variables

### **LEVELS OF MEASUREMENT**

- Measurement involves a procedure for finding the size, extent, volume or degree of something i.e the unit of analysis
- Measurement can be done at 4 levels
- 1.Nominal scale
- 2.Ordinal scale
- 3.Interval scale
- 4.Ratio/Continuous scale

### **NOMINAL SCALE**

- Lowest and least precise type of measurement
- It merely classifies elements into categories based on a given characteristic
- It dichotomizes elements into those that are similar
- It involves assigning numbers to the categories as labels or codes for researchers' convenience in analyzing data and communicating results
- Nominal scale does not allow for any form of mathematical operation
- Example if a researcher defines a nominal scale as
- 1.Female=1; Male=2; 2Female=/Male

### **NOMINAL SCALE CONT'D**

- Variables such as Sex (male, female),
- Religion (Christian, Moslem, Traditional),
- Marital Status (married, single, widowed, divorcee) and soon have categories as shown to which arbitrary names or labels are given without assumptions about relationships between categories.
- Nominal scales are useful only for identification purposes

### **ORDINAL SCALE**

- It's the next higher level of measurement above nominal scale
- It involves ordering of elements into categories on the basis of the degree to which they possess a certain characteristics
- This form of measurement is known as ranking.
- Example: categorizing families according to socio-economic status, e.g upper, middle, low and then rank in increasing or decreasing order
- Ordinal scale cannot supply information about the extent of difference between categories
- Scale merely tells us that A is greater than B, but cannot say to what extent
- Like nominal scale, mathematical relationships and operations are not possible.

### **INTERVAL SCALE**

- This is the next higher level of measurement to ordinal scale
- It combines the qualities of nominal and ordinal scales.
- It possesses an additional quality in the sense that equal distances between numbers represent equal distances in the variable being measured.
- Example-The Fahrenheit measure of temperature is an interval scale.
- The difference between 20F and 40 degrees Fahrenheit is equal to that between 80 and 100 degrees
- Limitation is the absence of a true absolute zero
- The point zero is set arbitrarily, there is no unique zero origin

### **RATIO OR CONTINUOUS SCALE**

- This is the most advanced or highest level of measurement
- Ratio scale combines the features of nominal, ordinal and interval scales
- Additionally it has the feature of absolute zero point, which signifies the complete absence of the characteristic or property
- The measurement scale allows us to compare scores by taking their ratios, because it is possible to locate an absolute or non-arbitrary zero point on the scale
- Ratio scale is a superior level of measurement, compared to others
- It gives the most precise interpretation of a variable.
- Examples-weight in kilograms, income in naira, height in meters

### **CRITERIA FOR SCIENTIFIC MEASUREMENT**

- Scientific method relies on sound and accurate measurement of variables
- Sound measurement must meet the requirement of validity and reliability
- VALIDITY refers to the extent to which an empirical measure or operational definition adequately reflects the true meaning of the concept under study.
- Validity is concerned with the question: Is the researcher measuring what he intended to measure?
- Does the operational definition truly reflect what the concept means?
- Validity refers to the degree to which operational definition or measuring scale truly represents the property or characteristic under investigation

### **RELIABILITY OF MEASUREMENT**

- Reliability refers to the degree to which a given measurement procedure will give the same description of that phenomenon if that measurement is repeated.

- It concerns whether a particular technique will yield the same result always if repeatedly applied to the same object
- The meaning of reliability is captured in the question: Do repeated applications of the operational definition produce consistent results under similar conditions?

#### • **BETWEEN VALIDITY AND RELIABILITY**

- Validity is the goodness of fit between operational definition and the concept it is purported to measure, while reliability is the consistency or stability of an operational definition or measuring scale.

#### **ERRORS IN MEASUREMENT**

- All measurement is subject to errors or imperfections
- Though a measurement technique should be precise, unambiguous and flawless, in practice it is difficult to achieve
- There are usually three sources of variations.
  1. True differences across units of analysis
  2. Biases introduced by measurement, and
  3. Errors of measurement due to random or chance factors

#### **RANDOM ERRORS**

- Random errors stem from:
  1. Object related factors-transitory swings in objects being measured
  2. Respondent caused problems (ignorance, fatigue, boredom, suspicion, negative attitude to questions)
  3. Investigator caused error: coding errors, data processing errors, faulty calculations,
  4. Environment of the respondent: respondent may be restrained in the presence of persons not trusted,
- FOREGOING shows that each measurement consists of 2 components: a true component and an error component.

#### **QUALITIES OF A GOOD RESEARCHER**

- A good researcher must have certain basic qualities and skills
- Amongst these are:
  1. Reading
  2. Listening
  3. Watching
  4. Choosing
  5. Questioning/Enquiry
  6. Summarising
  7. Organizing
  8. Writing/Communication
  9. Presenting and
  10. Reflecting

#### **OTHER MEANS OF ACQUIRING**

##### KNOWLEDGE

- Scientific mode is one out of many means of acquiring knowledge.
- It is the most reliable means to reliable knowledge.
- OTHER MEANS ARE –
  - Authoritarian Mode-knowledge sought from people who are socially or politically defined as qualified sources of knowledge.
  - Mystical Mode: knowledge is acquired from supernatural authorities such as prophets, divines, gods and mediums.
  - Rationalistic Mode: Acquire knowledge by strict compliance to the forms, tenets and rule of logic. Human mind can understand the world independently of observable or empirical phenomenon.

- This is the strategic approach used in conducting research
- It is the research plan or blue print
- Research design is the anchor of the scientific study
- It gives shape, form and identity to the research activity
- It answers the questions-what, where, when, how and by what means the research enquiry will be carried out

#### **4. RESEARCH DESIGNS**

##### **CHOICE OF RESEARCH PLAN**

- The choice of research design involves decisions relating to: what entities should be studied?- Individuals, groups, communities, organizations, nations?(Sampling unit)
- What aspects of characteristics of these entities should be studied? Income, Production, Consumption, Input use, Ethics, Performance, Health, Education, Socio-economics
- What period of time will the study cover or Time required?
- Where will the study be located?
- Where will the data be found?
- What type of data is needed?
- What kind of relationships will be anticipated/ studied?
- It embraces the methodology to employ and analytical technique

##### **COMPONENTS OF RESEARCH DESIGN**

- The sampling approach, the process of selecting units from population
- The observational approach involving the techniques of data collection
- The statistical design, analytical technique i.e the tools of analysis
- The operational design includes the procedures and techniques to carry out the sampling, observational plans and analyze data.

##### **TYOLOGY OF RESEARCH DESIGN**

- The basic classification of research is based on strategy (how you do it) and purpose (why you do it) of the research.

- Types of research can be grouped into these non-mutually exclusive categories.

1. Desk research
2. Field research
3. Laboratory
4. Pure
5. Applied
6. Exploratory
7. Descriptive
8. Explanatory/Analytical
9. Comparative Studies
10. Qualitative
11. Quantitative research

##### **DESK RESEARCH**

- Research is limited to library study and analysis of documents and/or the analysis of historical records
- It employs techniques such as recording of notes, content analysis, listening to films and tapes, compilations and abstract guides
- Research in social sciences involves library studies, review of secondary literature,, government records, diaries, calendars and historical documents.
- Most often, desk research is used in contradiction to research involving collection of primary data from respondents through interviews and questionnaires

### **FIELD RESEARCH**

- Involves going out to collect data
- Research involves observation using some techniques like- interviews, mailed questionnaire, focus group discussions, and key informant interviews.

### **• PURE RESEARCH**

- Comprises research regarded as theoretical or academic in purpose
- Refers to research aimed at testing conceptual hypothesis or developing theoretical principles
- Pure research entails doing research to gather knowledge for knowledge sake.

### **APPLIED RESEARCH**

- The basic feature is the empirical content. It's based on observation/experiment as against theory
- It refers to research that relies on observation with little regard for theory
- It is data-based and gives conclusions that are verifiable through observation or experimentation
- One variant is action research
- Action research is done for the purpose of resolving a particular problem. Its aimed at collection of information which can be immediately used to make practical decisions relating to development planning, program design and implementation • Such research concerns, diagnosis of a situation e.g what caused it, how are the factors related, if there is a stimulus, how does the situation change?

### **PURE VIS-A-VIS APPLIED RESEARCH**

- Pure and applied research though may differ in approach; there is interplay between them.
- Good theoretical research can be applicable to practical problems and practical/applied research can enable advancement in theory, e.g
- Applied research can contribute facts which may lead to the initiation, rejection and reformulation of theory, or its clarification and re-definition or re-examination.
- Applied research offers an opportunity to test the validity of existing theory and may aid in conceptual Clarification & precision

### **PURE VIS-A-VIS APPLIED RESEARCH**

- Applied research can be useful in integrating existing theory especially if and when the solution of a problem requires some integration of the finding from theoretical and applied investigations in diverse fields
- Pure research offers solutions to many practical problems by developing general principles. Although direct application of results is not an immediate objective of research, results of such research may eventually be widely applicable with far-reaching consequences.
- Pure research can go beyond common sense approach to abstract the key factors in a practical problem, thereby facilitating efficient solutions to the problem

### **PURE VIS-A-VIS APPLIED RESEARCH**

Pure research develops many alternative solutions to a particular problem, specifying main and auxiliary consequences, thereby permitting a choice of best solution. This is because pure or theoretical research extends beyond workable/feasible solutions to more precise generalizations, discovery of critical factors and ascertainment of exact conditions in question

### **EXPLORATORY RESEARCH**

- Exploration means to develop an initial approximate understanding of some phenomenon.
- Its applicable where relatively little/nothing is known about a process or where the situation has a deviant character or its relatively new
- Studies of exploratory nature are valuable because they can break new grounds, help to develop hypothesis and provide insights into the topic

- Exploratory research entails investigations that are used to obtain quick, not very precise information on development issues in local communities
- Rapid methods are used where time and resources are short but limited information is still of value.
- Rapid studies of this nature are called exploratory case studies, especially if they are on small-scale
- Typically, rapid methods involve a multidisciplinary team working in the study area, observing characteristics of interest
- Observation may be direct/indirect or by interviewing informants (local leaders, elders/those knowledgeable about the subject)
- Could be by group interview technique (group interview or focus group discussion)
- Exploratory study/rapid rural appraisal/reconnaissance survey/pilot study/preliminary study are other names.

### **DESCRIPTIVE RESEARCH**

- It refers to the identification and measurement of some characteristics of a population or phenomenon under study.
- A descriptive study entails the systematic collection and presentation of data to give a clear picture of a particular situation
- It is fact finding in nature.
- In descriptive research, the researcher is involved in *ex post* study, i.e. what has happened or is happening
- Descriptive studies employ case studies, sample surveys, census or aerial survey.

### **CASE STUDY**

- An enquiry in which limited cases or number of units are investigated in-depth.
- Unit selected may not be on random basis, but by possession of certain characteristics which are needed for the investigation, e.g. case study of unique people, episode or trend of development.
- We could find out whether the results of a case study pertain to large population
- For example case study focuses on observing the characteristics of an individual unit, a community, an institution, a school, prison, football team, a class, a work team and may later generalize to larger domains, where the unit belongs-This must however be treated with utmost care
- Case study is useful to illustrate good practice and reveal salient conditions. It allows for capture of complexity of social life and actual experiences of and real life situations.

### **SAMPLE SURVEY**

- This is an investigation in which only part or sample of the population is studied and the selection is made such that the sample is representative of the whole population
- Information gathered must be able to be generalized to the whole population.
- Sample survey can be cross-sectional or longitudinal in approach
- Quantitative sample surveys assess the distribution of established variables and measure their relationships cross-sectionally or longitudinally

### **SAMPLE SURVEY CONT'D**

- Cross sectional surveys aim at collecting information on certain variables in a study population at one point in time
- It may cover physical characteristics of people, knowledge, attitudes, beliefs and opinions, events and practices
- Longitudinal surveys involve observations over an extended period of time.
- There are 3 variants-Trend, Cohort studies and Panel studies
- Trend studies are concerned about changes within some general population over time
- Cohort studies focus on more specific subpopulations as they change over time
- Panel studies focus on the same sample over time
- Sample surveys are conducted by several agencies in Nigeria namely, National Bureau of Statistics (formerly FOS)
- Research Institutes, International Agencies(IITA), Non Governmental Agencies, Private Firms.

- For example, National Bureau of Statistics carries out the Nigerian Living Standard Survey (NLSS), the General Household Survey (GHS), Agricultural Sample Survey and Manufacturing Establishments Survey, Labour Survey, Consumer Expenditure Survey (CES), Core Welfare Indicator Questionnaire Survey (CWIQ) and many others

### **CENSUS**

- This is an enquiry which covers every individual unit in the population being studied.
- Example is the national censuses of population and of social phenomena which are conducted by many countries on regular basis
- However census may be defined for a smaller population e.g population of all farmers in a village, population of widows in a village, population of school children in a village.

### **AERIAL SURVEY**

- Survey involves an aircraft flying over the whole or part of an area of interest
- Information gathering can be by taking photographs or by an observer recording observations while in the air.
- Aerial survey is important for gathering information on topics which can be investigated by direct observation without interviewing or contacting respondent
- Examples are cattle and wildlife location, land use patterns, settlements
- Survey is very expensive in terms of money, manpower, skills and equipment

### **EXPLANATORY RESEARCH**

- Explanation involves finding out the reasons for a situation or phenomenon.
- It is essentially diagnostic
- It seeks to find out the frequency with which something takes place or frequency the event or process is associated with something else
- The key task is to find out –why e.g-why a rural area has higher infant mortality rates than others
- The term explanatory research is used interchangeably with comparative or analytical research
- Comparative studies compare 2 or more groups of things
- Comparison is a fundamental research strategy

### **CROSS-SECTIONAL COMPARATIVE STUDIES**

- Focuses on comparing and describing groups or phenomenon based on observation made at a point in time
- Example is a survey on malnutrition which attempts to establish percentage of malnourished children in a certain population, socio-economic, physical and other variables influencing food availability, feeding practices, the knowledge, beliefs and opinions influencing these practices
- In addition to describing these variables, the researcher can by comparing malnourished and well-nourished children, attempt to determine which socio-economic, behavioral and other independent variables contribute to malnutrition

### **CASE CONTROL STUDIES**

- Here investigator compares one group among whom a problem is present with another group called control where the problem is absent, to ascertain what factors have contributed to the problem
- For both cross-sectional comparative and case control studies, the researcher has to control for confounding or intervening variables
- Case control studies use stratification as well as matching to control for confounding variables
- Matching means ensuring that the cases and controls are similar with respect to the distribution of one or more of the potentially confounding variables
- Cross-sectional comparative studies and case-control studies are relatively quick and inexpensive

### **COHORT STUDIES**

- Have characteristics of both case-control and longitudinal study
- They are focused on examining more specific sub-population called cohorts, as they change over time
- Commonly used in public health research
- Example-enquiry where a group of individuals that are exposed to a risk factor is compared with a group of individuals not exposed to the risk factor, i.e control group
- Researcher monitors both groups over time and compares the occurrence of the problem expected to be related to the risk factor in the 2 groups so as to find out whether a greater proportion of those with the risk factor are actually affected
- For financial and practical reasons, cross-sectional comparative studies or case-control studies are preferred to cohort studies

### **EXPERIMENTAL STUDIES**

- Social research has 2 broad approaches-experimental and non experimental
- Experimental and non experimental social research differ in the degree of control they have over variables
- While experimental research manipulates variables suspected to produce effect, non-experimental research merely observes the variables and their effects
- In experimental research, the researcher applies a stimulus to a subject and observes the effect
- In experimental design, a researcher allocates randomly selected individuals to at least 2 groups. One group known as the experimental group is subjected to the intervention treatment while the other the control is not
- Both groups are investigated under conditions that are identical except for the treatment

### **EXPERIMENTAL STUDIES CONT'D**

- The condition is necessary to minimize variation between them
- The effect of treatment can be assessed by comparing the 2 groups
- The classical experimental study has 3 integral requirements manipulation or treatment or intervention, control and randomization
- Randomization means each subject is given an equal chance of being assigned to either group
- This allows holding all other factors constant so that differences between control and treatment can be linked to the applied treatment
- What is being done is to eliminate or minimize the effects of confounding variables
- Experimental research is common in pure sciences, education,
- Psychology e.g introduction of new curriculum

### **QUASI-EXPERIMENTAL STUDIES (QES)**

- A common QES uses 2 or more groups-
- One serves as control group and the other is the study group but the subjects in both groups have not been randomly selected or allocated
- Example- a researcher studies the effect of agric extension education on the level of participation of a village population in the adoption of new technologies
- Researcher selects one village in which extension education will be given and another that will not as control
- Survey is then done to find if adoption rate in one is significantly different from the other
- Study is quasi-experimental because subjects are not assigned on a random basis to either treatment or the control.

### **QUALITATIVE-QUANTITATIVE DIVIDE**

- A contemporary framework for classifying research in social sciences
- Definitional variants between the 2 can be linked to varying criteria which include nature of the subject under investigation, type of data required, the method of collecting data and nature of outcome of research



- Nature of subject under investigation-Qualitative research investigates the behaviour, motivations, desires, attitudes, beliefs, choices and opinions of people or groups of people. It is concerned with understanding actor's frame of mind

#### Nature of subject under investigation

- Its orientation is to describe and analyse culture and behaviour
- Example-Study to investigate utilisation of health services among women
- Its orientation is to describe and analyse culture and behaviour
- Example-Study to investigate utilisation of health services among women
- Quantitative research deals with measurable and quantifiable aspects of phenomenon being studied.
- Focuses on questions like to what extent? By how much? What relationship exists between factors?
- Concerned with phenomenon amenable to measurement of quantity and amount

#### Forms of data required/Method of data collection

- Qualitative research is concerned with collection and analysis of data in numeric form .
  - It involves data that can be subjected to rigorous quantitative analysis
  - -----
  - Method of data collection
  - Data collection methods used in qualitative research include group interview, focus group discussion, key informant interview, participant observation
  - Quantitative research data is based on controlled measurement that is verifiable and replicable.
- Instrument for data collection could be questionnaire.

#### Nature of Research

- Qualitative research is aimed at generating bases for theory or hypothesis
- In some cases it can be used for hypothesis testing and theories.
- Quantitative research takes a proactive stance in testing hypothesis and theories.
- In particular instances, quantitative research could lead to exploration of a topic for generating hypothesis

### **CHOOSING A RESEARCH DESIGN**

- Research designs are not mutually exclusive
- The chosen method should fit the research questions and the purpose of the research.
- A single research can combine different approaches and techniques. The dichotomy between desk and field research is unreal, since all researches involve desk research.
- When considering the choice of research design, researcher should select the approach to the enquiry, the scale of enquiry and the method to be used for data collection.
- It is possible that some research design may be more appropriate than others for investigation of different scales

### **TRIANGULATION PRINCIPLE IN CHOOSING RESEARCH METHODS**

- The process of using 2 or more methods to verify the validity of information being collected is known as triangulation
- The logic of triangulation is to use findings from one method, say quantitative investigation to check against a qualitative study, in a complimentary manner
- It could mean using a variety of information gathering methods by addressing the same issue using several different tools
- Research faces resource constraints in terms of availability of manpower, data processing facilities, finances and time. Investigation must be done bearing this in mind. Sampling is the selection of a number of study units from a predefined population or universe. All items in the field of inquiry make up the population or universe. The part from which information is collected is the sample

### **5 & 6 WEEK-SAMPLING**

### **WHY SAMPLE?**

- Population under study in some research may be small and so all study elements (units) may be included. However,
- Some study focus on large population, so it is neither feasible nor economic to obtain information from the whole population.
- Enumeration of a population (census) involves a great deal of time, money and resources usually beyond the reach of the researcher
- Even govt find it costly to conduct census, hence its often done once in a decade
- Besides, census may not give highest accuracy as many biased elements may increase with many numbers of observations
- A sample is justified based on cost and accuracy grounds.
- It is cheaper, faster and can provide sufficiently accurate results.
- Sampling permits economy of time, financial and scientific resources.
- In drawing a sample from population you address questions
- What group of study population is the sample to be drawn?
- How many people are needed (sample size)?
- How are these people to be selected (sampling method or technique to use)?

### **SAMPLING UNITS (STUDY UNITS)**

- A sampling unit is the object which the researcher intends to collect information, by either interview or measurement.
- It is the thing, person, event or phenomenon that constitutes the object of observation for the purpose of description, explanation and/or prediction
- The study problem determines/informs the study unit
- Sampling unit may be a geographical object such as a state, LGA, village, or social object such as household, family, individual, or enterprise, groups/association, plot of land, whole area cultivated,

### **SAMPLING FRAME**

- It is the list of every elements/member of the population from which the sample will be drawn
- In constructing the sampling frame, the definition of the unit of analysis is followed by the population of interest, which is the collection of the units that constitute the population
- Defining the population requires identifying the target population about which the results are to be generalized and constructing the sampling frame
- An example is the list of all households in a village
- Always ensure the frame cover the study universe

### **REQUIREMENTS OF GOOD SAMPLING**

- There 2 basic requirements
- A sample must be representative
- A sample must be adequate (of sufficient size)
- To be representative, a sample should provide a close approximation of the characteristics of the target population
- A representative sample allows for valid conclusions for the whole population
- It is wrong to make conclusions about a sub-section of a population unless it truly represents the population

### **ADEQUACY OF SAMPLES**

- A sample is adequate if it is of sufficient size to allow confidence in the stability of its characteristics
- In deciding the sample size, researcher is guided by the consideration whether the sample size is small enough to be affordable and manageable and yet large enough to give reasonable estimates of the population being measured

- A rule of thumb is if the basic requirement of representativeness has been met, we are more likely to get a better picture of the population, the larger the sample

### **SAMPLING METHODS**

- Sampling method or design depends on the research objectives and purpose, the accessibility of the population and availability of sampling frame
- Consideration of statistical properties of the different types of sample is a factor that impinges on the choice of sampling method
- The ability to estimate and reduce sampling error depends on the sampling method used
- There are 2 broad categories viz-Non probability and probability sampling methods
- Probability samples are oriented to be statistically representative while non-probability samples are theoretically or conceptually representative

### **EXAMPLES**

- NON-PROBABILITY Sampling methods include
  - Convenience sampling
  - Purposive sampling and
  - Quota sampling
- PROBABILITY Sampling methods include
  - Simple random sampling
  - Systematic sampling
  - Stratified sampling
  - Cluster sampling
  - Multistage sampling

### **NON-PROBABILITY SAMPLING METHOD**

- In situations when non-probability sampling methods are used, the sample is selected not by chance but for reasons of practical convenience, personal judgment, subjective inclusion or in order to include units with particular characteristics in the sample.
- Procedure should be used with caution and where probability sampling is not practicable because researcher cannot apply statistical theory to examine the properties of sample estimates

### **CONVENIENCE SAMPLING**

- It is non probability sampling method (NPSM) in which a researcher selects for the sample those units that happen to be available at the time of data collection those study units that he comes across
- Over-riding principle here are researchers convenience and respondent availability
- Risk of bias in the method is obvious
- For example, persons encountered in the study may not be representative of the villagers.
- Some characteristics of the villagers may be overrepresented or under-represented or completely left out

### **QUOTA SAMPLING**

- Method could yield more representative sample than convenience sampling.
- Quota sampling ensures that certain number of sample units from different categories with specific characteristics appear in the sample so that all these characteristics are represented
- There is restriction that certain characteristics of the respondent must be typical of the area or group being studied
- It establishes quotas for different groups in the population and by convenience or purposive sampling fill each quota
- Used only when quick and crude results will satisfy practical sense

### **PURPOSIVE SAMPLING**

- Method draws a sample to illustrate some particular characteristics in the population
- For example, selecting only those who have adopted a certain technology or practiced certain techniques
- A purposive sample can be based on the capacity of the respondent to supply the required information, as in pig farmers, rice farmers or poultry farmers
- A sample of areas or persons can be chosen purposively because they have certain typical characteristics
- Purposive sampling is weak because sample varies in unknown ways from the universe

### **PROBABILITY SAMPLING METHODS**

- It involves random selection to ensure that each unit of the sample is chosen on the basis of chance
- All units of the population should have an equal or at least a known chance of being included in the sample
- A probability sampling procedure requires a sample frame
- Probability sampling has the key advantage of minimizing risk of sampling bias and it enables a researcher to draw inferences, from the sample about the population with levels of confidence that can be estimated statistically

### **SIMPLE RANDOM SAMPLING**

- Sampling procedure which ensures that every unit in the study population has a known and equal chance of being included in the sample
- STEPS-1. Decide what is to be the unit of sampling?
- 2. Make a list of all the units in the population or use an existing sample frame
- 3. Select the required number of sampling units (sample size) using a table of random numbers or throwing dice
- Sample size depends on size of target population, variability of the population parameter which researcher wishes to estimate and the degrees of precision & confidence aimed at by the researcher

Minimum Sample size  $n$ , estimation

- $n = Z^2 S^2 / (e + Z S/N)$
- If the population proportion were to be estimated, the minimum sample size is
- $n = [Z P(1-P)]/e + [Z P(1-P)/N]$
- Where  $N$ = Population size,  $n$ = sample size
- $S$ = population variance of  $Y$  (assumed to be known)
- $P$ = population proportion of  $Y$  (assumed to be known)
- $E$ = desired level of precision
- $a$  =desired level of confidence (e.g. 95 %)
- $Z$  = distribution corresponding to a level of confidence

### **SYSTEMATIC RANDOM SAMPLING**

- A variant of simple random sampling
- It is characterised by a random start, followed by a predetermined or systematized order of selection
- Once the first unit has been selected randomly, all the rest of the units for the sample are predetermined
- PROCEDURE-1. List the population to be sampled
- 2. Divide total number of units in the sampling frame  $N$ , by the units of the sample  $n$ , to obtain sampling interval  $K$ , rounded off to the nearest whole number, e.g  $1400/275=5.1$  or 5
- First unit from population is selected randomly
- One random number between 1 and  $K$  inclusive is selected, say  $r$
- The rest units are chosen systematically by selecting every  $K$ th unit on the list
- Sample is made up of  $r, r+K, r+2K, \dots, r+(n-1)K$

- Method saves time and cost and retains features of probability sampling. It is easier and quicker than simple random sampling.
- Problem is when sample frame is not in random order e.g intrinsic regularity or periodicity that is population listed in order say rich to poor or large to small, selection process will be biased
- Where frame is ordered or ranked, then sampling frame would be satisfactory for use and can increase sampling efficiency.

### **STRATIFIED RANDOM SAMPLING**

- The purpose of stratified random sampling is to produce sub-populations where the units of each subpopulation are alike in one major characteristic
- Stratification is dividing study population into mutually exclusive sub-populations called strata and drawing a sample in a random manner
- Rationale for stratified sampling is that a homogenous population requires smaller sample than heterogeneous one
- Simple random sample may lead to wrong representation, stratification helps to achieve better representativeness
- Many variables or criteria may be used to divide the population into strata-eg age, income occupation
- Strata created must be more homogenous than the population, if not upheld then there is no advantage in stratification
- MECHANICS-1. Divide population into strata
- 2.Consider each sub-group as a separate population, draw sample using some random type procedure such as proportional, equal or optimum allocation sampling
- PROPORTIONAL sampling means an equal percentage of each stratum is selected, regardless of stratum size
- Equal-size sampling means an equal number of units is randomly selected from each stratum, eg selecting 50 units from 3 strata to have 150 sample size
- There has to be adjustment for varying proportions in the separate strata by use of a weighing factor. Weight for under-represented group is
- $WEIGHT-W = P*N/(n(1-P))$
- Where W= weight,
- P= correct proportion of the under-represented group in population, n=no of under-represented group in the sample
- N= no or size of the rest of the sample, the over-wt group
- If no weighting is done, the data should be presented separately for each group and not pooled
- OPTIMUM allocation sampling involves using formula to determine the sample size for each stratum that will yield the most precise estimates
- Also it is necessary to adjust for varying proportions of the sample in the separate strata, using weighing factor

### **CLUSTER SAMPLING**

- Sampling method divides the population into contiguous groups called clusters
- Samples are then drawn from the clusters by random method
- Clusters are often geographical areas/units( e.g villages districts ) or organizational units( farmers cooperative society, projects clinics)
- Clusters can also be groups of households or animals or herd
- **Cluster sampling is the selection of groups of study units** (clusters) instead of study units individually
- Where there is no complete frame, researcher can compile group lists and a random selection done from group lists

### **CLUSTER SAMPLING MECHANISM**

1. Construct a sampling frame of the clusters, villages, industries and so on
  2. Randomly select clusters to be included in the study
- Number of clusters depends upon the needs of the research

- 3 for each of the clusters included draw the units of sampling in a random manner
- This may mean constructing sample frame for each selected clusters and use simple or systematic random sampling method to draw the sample from each cluster

### **MULTI-STAGE SAMPLING**

- Method of sampling involves procedure where the selection of units into the sample is organized into stages
- It usually involves a combination of sampling methods
- There can be 2,3 or many or multi-stages involved
- In say 3 stage sampling a researcher may draw a sample of say districts within a region (first stage), villages within districts (second stage) and lastly a sample of households within each selected villages (third stage)
- Attribute of multi stage sampling is that it is possible to reach final sampling or study units through a hierarchy of higher level stages

### **SAMPLING METHOD AND SAMPLE SIZE**

- Avoid wrong premises based on unfounded stereotypes, e.g.
- 1. That probability sampling is associated with large scale studies and non probability sampling with small scale study
- 2. That probability sampling is used together with formal questionnaire and non probability sampling with unstructured interview
- There is no justification to presume that a single study must involve either probability or non-probability sampling
- In many practical cases a single study combines both sampling procedures

### **CRITICAL CRITERIA IN DETERMINING SAMPLE SIZE**

- Factors that should be considered are:
- The level of precision required in the estimates
- The confidence level
- Level of variability of the factors (variables) to be estimated
- The sampling method used
- The homogeneity of the population
- The prior information about the population
- The number of categories of data to be analyzed
- The level of disaggregation of the study results
- Resource constraints

### **Level of Precision**

- Level of precision is also called the sampling error
- It refers to the range expressed in percentage points, eg 5 percent, within which the true population value may be found
- The level of precision is the closeness with which the sample predicts where the true values in the population lie
- The difference between the sample and the real population is called sampling error
- If a study gives result that 70% of students are using a service with precision rate of  $\pm 5\%$ , then it can be concluded that between 65% and 75% of the students population are using the service
- The more precise the estimates desired, the smaller the standard error and the larger the sample necessary to obtain a given level of precision
- Higher levels of precision require larger sample sizes and higher costs to achieve
- CONFIDENCE LEVEL-denotes the risk which the researcher is willing to accept that the sample is within the average or bell curve of the population

- A confidence level of 90% means 90 out of 100 repeated samples of the population would have the true population value within the range of predefined precision while the remaining 10 would be unrepresentative samples
- Put in another way, a study may posit that the estimate would be precise within  $\pm 10\%$ , given 95% confidence level
- This means that an error greater than 10% points would only occur in not more than 5% of all possible samples
- The chosen confidence level is expressed as a value of Z statistic, from the normal distribution
- If sample size  $n > 30$ , then the values of Z are 1.96 at 95% and 1.64 at 90% and 2.58 at 99%
- For small samples  $n < 30$ , the confidence level is derived from t-distribution and the values are larger than shown for Z

#### **THE LEVEL OF VARIATION OF POPULATION ATTRIBUTE**

- The degree of variability is the extent to which the attributes being measured/observed is distributed or dispersed throughout the population
- In order to reach a given level of precision/accuracy, every study requires larger sample for more heterogeneous population and smaller sample for relatively homogeneous population

#### **SAMPLING ERROR AND STANDARD ERROR**

- Sampling error is the imprecision that results due to the fact that an enquiry has been limited to a part of the population (sample) rather than the entire population
- A difference exists between estimate and the true population value and is due to the use of sample
- This element of difference is known as the sampling error
- Error here is in strictest sense a technical terminology used to denote uncertainty introduced by sampling
- Every sampling is subject to sampling error
- The standard error indicates how close the estimates from a single sample is likely to be to the true population value
- Statistically, the standard error is the standard deviation of the sampling distribution of means
- While standard deviation measures dispersion about an actual distribution, standard error measures the dispersion, about the mean, of a hypothetical distribution
- Given that there exists a relationship between the standard error of a sample mean,  $S.E(\bar{X})$ , and the standard deviation of the population observation, it is  $S.E(\bar{X}) = S.D(X)/\sqrt{n}$
- The standard error of mean population is the standard deviation of population variable divided by the square root of sample size

#### **NON SAMPLING ERRORS**

- Non sampling error covers all kinds of errors which arise even if the whole universe is studied:
- It includes human mistakes and wrongs which span the research duration
- Measurement errors
- Error due to improper sample selection
- Non response errors
- Interviewer's own biases

#### **CONTROLLING NONSAMPLING ERROR**

- Proper sampling procedure
- Pre-testing of data collection instrument
- Follow-up of non response if due to absence
- Include additional people in case of continued absence
- Carefully construct the data collection instrument, subject to construct validity and content validity tests
- Ensure sample frame is complete and up-to-date
- Check data entry, coding and processing for verification and validation purposes

#### **SAMPLING ERROR AND SAMPLE SIZE**

- The total error is a function of 2 independent sources of error and cannot be significantly reduced unless both types of errors are simultaneously controlled
- If non-sampling errors are large, there is no use taking a large sample so as to reduce the standard error of the estimate, since the total error will be determined by the length of the base of the triangle (ie the non sampling error)
- Likewise, if non sampling errors were to be brought to minimum, there is no point taking small sample, thereby having a large standard error
- Maintain a proper balance between sampling and non sampling errors total error =
- Sampling error
- + non-sampling error Its concerned with estimation of parameters and testing of hypotheses for significance

## 7. STATISTICAL INFERENCE AND TEST OF HYPOTHESIS

- Statistical inference means drawing conclusions about the population by studying a sample of the population
- Validity of statistical inference is based on the theory that values in a population are normally distributed and that the statistics from a large enough random sample of that population will be randomly distributed
- Statistical inference works on the assumption of specific population and sampling distribution
- Its important to understand the common distributions viz normal, t-distribution, F-distribution, Chi-Square distribution

### NORMAL DISTRIBUTION

- Normal distribution is a probability curve which is bell shaped and symmetrical about the mean
- It's a continous distribution which describes the shape and form of the probability density function of a variable
- For a variable X with normal distribution, mean  $\mu$  and probability that X takes a value  $X_1$ , that is  $p(X_1)$ , the normal curve for the variable is given as:

### AREA UNDER THE CURVE

- The area under the curve is measured in units of standard deviation so that the normal curve expresses the probability of any value occurring as a deviation from the true value (the mean)
- The normal curve has a unique property that allows us to obtain the proportion of cases falling within a given interval
- Based on this property, there is a constant area between the mean and an ordinate which is a given distance from the mean, in terms of standard deviation units

### NORMAL DISTRIBUTION

- Each portion of the curve represents some portion of the total area, which is one
- If we extend one standard deviation to the right and to the left of the mean, we shall have  $.3413 + .3413 = .6826$  cases included.
- The area between the mean and 2 standard deviations away on either side of the mean ie ( $\pm 2SD$ ), includes  $.4773 + .4773 = 95.46\%$  of the total area.
- All cases are noted to be practically within ( $\pm 3SD$ ) away from the mean.
- It must be noted however, that normal curve extends indefinitely far on either side of the mean
- Note that distance from the mean can be fractions and not exact multiples of the standard deviation.
- For example, if we extend ordinates that are 1.96 SD on either side of the mean, we get exactly 95% of the cases
- We get 99% of the cases if ordinates are 2.58SD away from the mean on either side
- Normal distributions lack standardization, thus complicating the computation of probabilities



- We therefore standardize variable X by transformation into standardize normal distribution Z, with mean zero and unit variance as follows  $Z_i = (X_i - \mu)/SD$ ; and  $Z = (\bar{X} - \mu)/(SD/\sqrt{n})$
- $Z \sim N(0,1)$ . Variable Z is measured in SD from the origin

### STANDARDIZED NORMAL DISTRIBUTION

- It is the property of a standard normal curve that the probability of Z falling between  $\pm 1SD$  is 0.6825
- The probability of Z falling between  $\pm 1.96SD$  is 0.95
- The probability of Z falling between  $\pm 2SD$  is 0.9544
- The probability of Z falling between  $\pm 3SD$  is 0.9974
- Statistical inference using the Z transformation is predicated on: 1. The population variance is known, no matter size n, or
- 2. Population variance is unknown and sample size,  $n > 30$
- Alternative conditions are based on the fact that when sample size is large, sample variance approaches population variance

### ILLUSTRATION

- Assume that a normal distribution has a mean of 168 and SD of 12.
- We can find the proportion of the curve indicated by Z representing the range from the values 143 to 168. Therefore,  $Z = (143 - 168)/12 = -25/12 = -2.08$
- The negative sign means that Z is located to the left of the mean, as shown
- From Z-table, the area between the mean and a Z value of 2.08 is 0.4812
- Since the total area is unity and area to the left is 0.50, the shaded area is  $0.5 - 0.4812 = 0.0188$
- If we are interested in the total area outside the region, then there will be 2 equal shaded areas, shown as:
- When to apply the relation is answered by the central limit theorem which states that if sample size is large, sampling distribution of the means would be normal whether or not the distribution of population of X is normal, ie.  $n > 30$

### The Students t-Distribution

- If population variance is unknown, as is usually the case and sample size is small, ( $n < 30$ ), a transformation based on student t-distribution is done, provided that the population of the variable under study is normally distributed
- This is because the sampling distribution will no longer be normally distributed
- Instead of assuming a normal distribution of sample means and using a Z-statistic, we use student t distribution, computed as  $t = (\bar{X} - \mu) / (SD/\sqrt{n})$
- Values of t-distribution at various df are tabulated

### TESTS OF HYPOTHESIS

- Test of significance involves testing the difference between statistics from 2 or more groups within the sample
- Statistical tests are done because the fact that one estimate is larger than the other does not imply that there is surely a difference-the difference could be due to chance
- The significance test will tell us the probability that the difference between the 2 estimates could have occurred by chance
- The procedure is called test of hypothesis-Null vs Alternative
- The assumption made which may or may not be true is made about the population
- The null hypothesis denoted by  $H_0$  is the one which we want to reject

### NULL HYPOTHESIS

- Null hypothesis says, that any observe difference between 2 estimates is due to chance ie the sampling error
- The hypothesis which differs from the null is called the alternative hypothesis,  $H_1$

- LEVEL OF SIGNIFICANCE, TYPE I and II ERRORS
- Level of significance and error of judgment are critical issues in hypothesis testing
- If we reject a hypothesis when it should be accepted, we commit type I error-implying taking incorrect action.
- If we accept a hypothesis when it should be rejected, we make a type II error
- In testing hypothesis, the maximum probability with which we would be willing to risk a type I error is called the level of significance of the test
- It is the likelihood that a null hypothesis which in fact is true will be rejected
- A level of significance of 0.05 and 0.01 are common
- 5% level of significance means there are 5 chances in 100 that a researcher will reject the null hypothesis when it should be accepted.
- That means 95% confidence that the right decision has been made. There is 5% probability that the researcher will be wrong
- Minimizing error of judgment is important for hypothesis testing
- Note that an attempt to minimize one error increases the other error type.
- But committing one error may be serious than another.
- Compromise should be made in limiting the serious error
- In general, probability of type I error is larger than type II error because type I error is what the researcher is avoiding
- Raising level of significance increases the chance of type II error, lowering it increases the chance of type I error.
- The only way to reduce both types is to increase sample size

#### **One-tail and Two-tail test**

- The null hypothesis  $H_0 : \mu = 230$  is a two tail test
- $H_0 : \mu > 230$  is a one tail test
- GRAPH of two tail test

To develop data collection plan, consider these questions: What information is needed to answer the questions implied by the research objectives?

#### **8&9 DATA COLLECTION**

- From whom will the information be collected?
- How long will it take to collect the data?
- In what sequence should the data be collected?
- What most suitable methods can be used to collect the data?
- How can the quality (reliability and validity) of data be ensured?
- What time and budget limitations exist that can impinge upon the data collection activity?
- These basic questions can be answered within the context of research problems, study objectives, stated hypothesis and use of result.

#### **TYPES OF DATA**

- In every research there are 2 types-primary and secondary data
- Primary data are those collected afresh, for the first time by the researcher
- Secondary data refers to those which have been already collected by someone else.
- Can be obtained from existing literature, research reports, government documents, Institutions publications and Statistical Reports.
- Using secondary information may pose problem to researcher because its not collected for that purpose

#### **QUESTIONS TO POSE IN USING SECONDARY DATA**

- Are the data too old?
- Do the data bear the frequency and timing required?
- Do data have enough information on relevant research variables?
- Do data cover geographical space?

- Are data in the forms and types that are useful to the study?
- Is the sampling that produced the data compatible to study?
- Is the data aggregation compatible?
- What is the integrity of the data collection process?

#### **TYPES OF DATA FROM SURVEYS**

- Research data can also be classified according to
  - TIME DIMENSION
  - DEGREE OF AGGREGATION
- According to time dimension, we have cross-sectional, time series and panel data (longitudinal)
- SINGLE Cross- section data are facts on research variables at a given point in time.
- They have been generated by a slice or cross-section of the population at a point in time.
- Disadvantage is the absences of data from more than one point in time- no inter-temporal variability.

#### **DATA FROM REPEATED CROSSSECTION**

- Situations where cross-sections are taken at regular intervals
- Where the repeated cross sections are drawn from the same sampling frame, the cluster panels are created
- This permits analysis of inter-temporal variation within the cluster
- Repeated cross-sections are NOT panel surveys because they do not return to the same sample even though covers same population
- ADVANTAGE-Able to capture inter-temporal variation unlike single cross-section

#### **TIME SERIES DATA**

- It gives information about research variables from period to period
- Called time series because variables are measured at regular points of time eg Annual GDP/GNP, National birth rates
- **PANEL DATA** –give values of the same set of variables over time, arising from repeated surveys of the same unchanged sample
- Respondent is interviewed at regular intervals to update information
- ADV-Gives estimation of changes in variables
- DISADV-May suffer from respondent fatigue leading to non- response or drop out

#### **DATA CLASSIFICATION ACCORDING TO DEGREE OF AGGREGATION**

- MICRO-LEVEL data, measure variables within the behavioural units, itself eg. Household or firm/enterprise
- MACRO-LEVEL data –measure behaviour for a group of such behavioural units.
- When data is observed for households it is called micro-data, while data on whole economy say, GDP/GNP. Employment and population are macro or aggregate data

#### **DATA COLLECTION METHODS**

- Generally 2 approaches
- 1. DIRECT APPROACH involving measurement or observation
- 2. INDIRECT APPROACH involving interviewing respondents, distribution or administration of questionnaire
  - 1. Direct Approach- a) Physical measurement by researcher
    - b) Observation– non-participant and participant methods
  - 2. Indirect Approach— a) in depth interview
    - b) Focus group discussion
    - c) Group interview
    - d) Individual interview
    - e) Formal(mailed) questionnaire

Data collection by measurement

- Measurement is the process of assigning numbers or labels to units of analysis.
- A compromise approach between physical measurement and interview method involves collecting required data by interview from the entire sample and also measurement from a sub-sample
- Comparison of the 2 sets is then done to see level of accuracy

#### Data collection by Physical Observation

- It involves live witnessing and recording of the characteristics of an event, occurrences or persons as defined by research
- It can give more accurate information than interviews or questionnaire
- Observation takes place when we see, hear, touch, taste or smell something.
- Observational approach is associated with varying levels of involvement of researcher or observer

#### Participant/Non participant Observation

- Participant observation is when the researcher or enumerator takes active part in the situation or activity under observation
- Non participant observation takes place when a researcher or enumerator watches the situation or people engaged in an activity and records the relevant aspects of their activities
- Examples are observing and recording the number of hours worked for, household consumption of commodities

#### Data collection by Interview Method

- There are situations when information may be sought by interview rather than measurement or observation
- Information may be measurable but requires much time and funds.
- Also it could be a case where information cannot be directly observed because of practical reason
- It may be information about peoples' knowledge, opinions, attitudes, beliefs, perceptions or values
- These situation necessitate the use of interview to gather information

### **INTERVIEW METHOD**

- An interview is a data collection method that involves oral questioning of respondents, either individually or a group
- The interviewer records the answers to the questions posed by writing them down or tape record them
- Interviewing connotes the development of precision, focus, reliability and validity in the act of conversation
- The researcher must develop interviewing skills
- This means learning to avoid certain types of error, alert to ambiguities, recognition and control of interview bias

#### Conducting an Interview

- Interviewer needs to establish rapport with the respondents, meaning respondent accepting research goals and willing to cooperate in answering questions
- Researcher gives explanation about the purpose of the research and assures confidentiality of information supplied
- Circumstance of interview may be a crowd of family members or neighbors. This may be an advantage in remembering answers and may be disadvantageous for confidentiality. Necessary to match circumstance

#### Interview Structure and Content

- Researcher needs to decide whether the interview will be
  1. unstructured and free-ranging
  2. structured and restrictive
  3. whether interview content will be closed ended or open –ended questions
  4. a combination of both

- Unstructured interviews are free ranging with flexibility and generally employs interview outline called interview guide
- The questions are open-ended since respondents are unrestricted in answers
- Structured interviews show low degree of flexibility by using questionnaire or interview schedule ie a fixed list of questions in a standard sequence
- Questions are closed-ended
- For large scale sample survey, open-ended questions may not be suitable

#### Reference and Recall Periods, Interview Length and Frequency

- Reference period is the time in the past during which the event being studied took place
- Recall period refers to the length of time into the past which a respondent is asked to remember an information
- Both reference and recall periods affect the quality of data
- Memory recalls can be subject to errors, inconsistencies, biases and memory lapses
- Ideal length of recall period will vary according to significance of event and the frequency of its occurrence
- A few days reference period is appropriate for frequent events like household food items purchase, durables 6- 12mon

#### Data collection by In-depth Interview

- Where in-depth interview is needed to explore research ideas or hypothesis, it may not be enough to narrowly define the topic
- It is best to conduct an in-depth interview where respondent feel most convenient
- Since essence is to get detailed information interviewer should encourage respondent to discuss freely
- Respondent for in-depth interviews are selected to suit pre-defined characteristics eg. Location, employment

#### Data collection by key Informant Technique

- Key informants are persons considered experts with special knowledge or insights of interest to the researcher
- Respondents of in-depth interview will not be expected to have any special knowledge. They just provide information in terms of their own personal experience
- A key informant is a person who is especially knowledgeable on topics of research interest
- Key informants interviews are combined with other methods of data collection

#### Data collection by Focus Group Discussion

- Method involves a small group conversation which allows the researcher to gain insight into the target population's knowledge, attitude and behavior pertaining to a certain topic
- Focus groups are particularly important in answering questions of how, what and why?
- Group dynamics needed are-homogenous with respect to certain characteristics of interest to the research
- Facilitator should be flexible and encourage contribution
- Limit discussion to specific issues

#### Data collection by Group Interview (Semi-Structured Interviews)

- Group discussion is not the only form of group interview
- Other kinds of group interviews can be conducted in group settings
- The group interview technique is a semi-structured question and answer discussion of a given topic by a homogeneous group of say 6-10 people
- It is not as more controlled and structured as focus group discussion.
- To conduct a group interview, researcher should do these-

To conduct group interview, Do these

- Identify target groups to interview
- Develop a group interview guide
- Determine the number of group interviews and the number of groups of each type to be interviewed
- Pre-test and revise the interview guide
- Determine the number of participants per group
- Determine the number and location of interview sites
- Conduct the group interview

#### Example of Group Interview Guide on Family Planning

- Target group: Women 20-35 yrs, married with children
- What is the value of children to people you know?
- Do families have a preference for girls or boys?
- If yes, why does this preference exist?
- When people talk about having many or few children, what do they mean by many and few?
- What are the advantages of having few children?

#### Disadvantages?

- Do men prefer many too few children? Why?
- Is there any number of children women prefer to have?
- Does religion encourage people to have many or few/
- What is the usual time period between births/

#### DATA COLLECTION: Questionnaire or Interview Schedule or Checklist?

- A QUESTIONNAIRE IS NOT THE SAME THING AS AN INTERVIEW SCHEDULE-though commonly used interchangeably
- Questionnaire is a device for obtaining answers to questions by using a researcher- prepared form, which the respondent fills himself
- Interview schedule refers to a form containing a set of questions which are asked and filled-in by an interviewer in a face-to- face encounter with the respondent
- An interview guide is merely a list of topics which an interviewer must cover during the interview

#### Questionnaire / Interview Schedule/Checklist

- Questionnaire is respondent administered, while interview schedule is researcher or interviewer administered
- Both techniques imply formal interviews
- Questionnaire can be likened to a written interview while an interview schedule resembles an oral interview.
- Researcher fills in the interview schedule, following oral responses given by the respondent
- The 3 data collection instruments comprise a set of related questions all logically related to the central problem, but they differ only in form, degree and structure
- Questionnaire and interview guide employ mainly structured means, interview guide use a greater proportion of open-ended Q
- Questionnaire is not feasible where respondents are illiterates

#### Designing data collection instrument- Questionnaire

- 1.Right questionnaire is crucial to the quality of survey
- Developing the instrument is effort -intensive
- What are the important questions?
- How are they to be asked?
- Researcher needs to consult other scientists, literature, carry out pilot study, involving Rapid Appraisal
- Well executed and adequately utilized pilot study will constitute a logical guide upon which the construction of the instrument will be based
- 2.Study the instruments used in related previous study

### 3. Write questions to be included

- This is the wording stage. Can have dramatic effect on response
- General principle is relevance, legibility, and accuracy
- Relevant questions match the purpose of the study
- Legible questions are those that read well
- Accurate questions are those which collect needed information in reliable and valid manner
- STYLE and TYPE of questions also influence responses
- Will the respondent be able to **understand** the question?
- Will the respondent be able to **answer** the question?
- This concern is important because a well understood question will elicit accurate answer

### **BOSS PRINCIPLE**

- BE BRIEF: Questions should be short because long questions become complex and confusing
- Recommendation is - a question should not be more than 20 words and no more than 3 commas
- Brevity should not be at the expense of losing meaning and question effectiveness
- BE OBJECTIVE-Objective questions are those which are neutral and do not prompt for a particular answer
- Non-objective questions bias the mind of respondents
- Questions should not suggest that a particular answer is the correct one eg. How effective is govt tax office in administering taxes and levies? With options-v effective, slightly, not so effective?

### **BE SIMPLE / BE SPECIFIC**

- BE SIMPLE: Use simple words, concepts and phrases in clear language and don't use technical terminologies
- BE SPECIFIC-Ask precise questions.
- When questions are vague, answers will be vague too.
- Avoid general, complex and ambiguous words
- A question should not attempt to obtain too much data at a time, that is loaded beyond what respondent can comprehend
- Avoid in-definitive words like often, occasional, regularly, rarely, generally, normally, good, bad, like, dislike
- Rather use definitive words –never, always, ever,
- In addition to observing principles of BOSS, note these

### **USE OF OPEN ENDED AND CLOSE ENDED QUESTIONS**

- Open ended questions allow respondents to express themselves
- There are no predetermined answers
- Main disadvantage is they take more time in analyzing and summarizing given the different respondents and many responses
- Makes analysis labour intensive
- There are circumstances when they are necessary despite these disadvantages eg where knowledge of respondent is being measured, when doing reconnaissance or exploratory study

### **USE OF SUBJECTIVE QUESTIONS**

- Subjective questions are those that measure peoples'opinion, knowledge, feelings and perceptions, ie opinion questions
- Subjective phenomenon is known only by the person. Responses are unique to the person. No means of independent verification. No objective standard to measure its wrongness
- Example-There is no guarantee that "agree" from one respondent is different from "strongly agree" from another
- What a respondent perceive as good may be considered as fair or even poor by another
- **Qualitative questions** entail use of rating scale for answers
- Principles to observe are-number of categories, which may differ

- If a neutral answer is likely, there should be a midpoint or middle category eg in the question-Do you agree or disagree with; government is providing good water supply services?
- Response category are-strongly agree, agree, neither agree nor disagree, disagree, strongly disagree
- Scale categories should be sharp to reflect respondents' opinion in unambiguous manner
- In assigning numbers of rating scale, researcher can use rank ordering. Question can contain a list of responses and respondent is asked to rank in order of importance, from most to the least important. Number can be assigned as –not important=1, important=2, very important=3. Or: poor/fair/good

#### 4. Arrange questions in logical sequence

- Important because it eases the interview process
- Arouses and retains respondents interest
- Overcomes respondents suspicions or doubts
- Aids recall of information by respondents
- Motivates respondents to collaborate

#### • 5.KEEP REALISTIC LENGTH AND FINE LAYOUT

- 6. CARRY OUT A PRE-TEST-This is rehearsal of the survey instrument
- It is the trial administration of the instrument among audience
- Motive is to ensure that the Question design is appropriate
- Pre-test is a verification tool to find out if Qs have same meaning
- Size of pre-test sample depends on resources and logistics

#### PRE-TEST GOALS

- To evaluate the suitability and adequacy of the questionnaire
- To assess the length of interview
- To test the quality of interviewers
- To assess the analytical value of the questions
- Through pre-test researcher identifies a number of design defects that should be corrected
- Defects commonly revealed include
- Problems with reading
- Problems with instructions
- Problems with questions clarity

#### PRE-TESTING

- Problems with assumptions about respondents-  
Recall period
- Problems with sensitivity- uncomfortable with question
- Problems with response categories-some categories of response overlap, some are ambiguous and not easily discerned
- There could be cases of missing responses, ie suggested answers are not exhaustive
- 6.Revise and finalize the instrument

#### DATA PROCESSING AND QUALITY CONTROL

- It involves the handling of data from one medium to another
- The tasks involves
- Data sorting-necessary where the study has different populations, data is sorted into groups.
- Could number Questionnaire
- Categorizing of Data-Necessary where there is use of open ended questionnaire. Decide how to categorize
- Coding of Data-Involves translation of all responses into identification numbers or symbols prior to analysis
- Data transfer to Master Sheets-Broad sheets in manual analysis



181

#### DATA TRANSFER USING COMPUTER SOFTWARE

- After coding and checking data for transcription error, data is ready to be entered into computer
- Prior to this design data entry format which gives the fields where data is entered is done
- Following data entry and storage, the data is manipulated to working files preparatory to data analysis

#### DATA QUALITY CONTROL

- This is crucial to the credibility and integrity of research
- There is need for Data Verification and Data Validation
- Verification examines for completeness and consistency
- Validation finds whether data satisfy criteria of magnitude & logic

#### SOFTWARE FOR DATA ANALYSIS

##### • S/No

##### • Economic and Statistical Procedures

##### • Common software

- 1
- Univariate analysis-central tendency and dispersion
- Microsoft Excel,SPSS,SAS,STATA,etc
- 2
- Bivariate analysis-Correlation and Regression
- MicrosoftExcel,SPSS,SAS,STATA,Econometric Views
- 3
- Inferential Statistics-t=test,ANOVA,Regression
- MicrosoftExcel,SPSS,SAS,STATA,LIMDEP
- 4
- Inferential Statistics-Chi-Square,Normal Distribution
- SPSS,STATA,SAS,LIMDEP
- 5
- Models of Qualitative Choice-Probit,Logit,Tobit
- SAS, STATA,SPSS,E-Views

Researchers use variables to denote characteristics of units that differ from one unit to the other. Variables may vary over units, time or over both Data analysis serves as operational tool to find out how variables are related based on empirical observation

#### 10 &11 DATA ANALYSIS

##### QUANTITATIVE VARIABLES

- Analysis of quantitative variables vary depending on the number of variables involved
- Analysis of quantitative data can be categorized as follows;
- Univariate Analysis/ Bivariate Analysis /Multivariate Analysis
- UNIVARIATE analysis involves the values of a single variable
- It seeks to find out the behavior of the variable in terms of tendency or pattern of values
- MEASURES OF CENTRAL TENDENCY
- Central tendency of a distribution is an estimate of the center of a distribution of values around the mean or midpoint
- Three major types are-MEAN, MEDIAN, MODE

##### THE MEAN

- Most common measure of central tendency is the mean
- It is the sum of observations divided by the number of observations
- Mathematically,  $\bar{X} = \frac{\sum X_i}{n}$
- Where  $X_i$  is the  $i$ th observation of the variable  $X$ ,
- $\sum X_i$  is the summation of values of all  $i$ th observations of  $X$

- N is the number of observations of X
- X is the mean
- For GROUPED observations  $\Sigma X = \sum f_i m_i / \sum f_i$  where
- $f_i$  =number of cases in ith category with  $\sum f_i =N$
- $m_i$  = midpoint of ith category, k = no of categories

**MEDIAN**

- It is the score found at the exact middle of the set of values
- To find, rank observations from low to high, locate the score in the middle of the sample
- If N is even, there will be no middle value; median is computed by averaging the two middle values
- MODE is the most frequently occurring value in a series
- MEASURES OF DISPERSION
- Dispersion is the tendency of values of a variable to scatter away from the mean or midpoint.
- Measures of dispersion include range, mean deviation, quartile deviation, variance , standard deviation and coefficient of variation=standard deviation/Mean

**RANGE**

- Range is the difference between the highest and the least value
- The quartile is a type of range and is computed as half the distance between 1st and 3rd quartiles ie  
 $Q=(Q3-Q1)$
- It is focused on variability among middle cases, and ignores the extremes of the distribution
- MEAN DEVIATION is defined as the arithmetic mean of the absolute differences of each observation from the mean.
- Computed as Mean Deviation= $(\sum |X_i - \bar{X}|) / N$

**VARIANCE**

- VARIANCE is the arithmetic mean squared deviations from the mean i.e  $s^2$
- $s^2 = \sum (X_i - \bar{X})^2 / N$
- Standard deviation is the square root of the variance,
- Computed as  $s = \sqrt{\sum (X_i - \bar{X})^2 / N}$
- COEFFICIENT OF VARIATION,  $CV=(s/\bar{X})$
- OTHER Univariate Analysis Techniques are
- Frequency Distribution
- Proportions. Ratios ie a/b and Percentages= $a/b * 100$

**BIVARIATE ANALYSIS**

- Bivariate analysis is concerned with the relationships between pairs of variables (X,Y) in a data set
- It is the simultaneous analysis of the two variables, usually undertaken to find out if one variable is related to another variable
- Examples of bivariate analysis include
- Cross tabulation, Simple correlation and Simple linear regression
- When 2 variables are classified jointly, the resulting table is known as cross-tabulation, 2 way cross tabulation

**EXAMPLES OF CROSS TABULATION**

- Table Illustration of Cross Tabulation

Respondent no of children	Male	Female	Total
One	43 (19.1)	26 (11.3)	69 (15.2)
Two	67 (29.8)	89 (38.9)	156 (34.4)
Three	71 (31.6)	65 (28.4)	136 (30.0)
Four	32 (14.2)	43 (18.8)	75 (16.5)

Five or more	12 (5.3)	6 (2.6)	18 (3.9)
TOTAL	225 (100)	229 (100)	454 (100)

**SIMPLE CORRELATION ANALYSIS**

- It attempts to find out the degree to which 2 variables are associated or tend to move together
- When the association is between 2 variables it simple correlation.
- When there are more than 2 variables, it is called multiple correlation. Possible correlations are:
  - a.Positive Linear Correlation
  - b.Positive non linear correlatn
  - c.Negative linear correlation
  - d.Negative non linear correlation
  - e.Zero correlation

Measuring linear correlation

- For a precise quantitative measurement of the degree of correlation between X and Y variables, we use a parameter called the correlation coefficient- $r_{xy}$ , measured using a sample by  $r_{xy}$
- $r_{xy} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}}$
- Where  $x_i = X_i - \bar{X}$  ; and  $y_i = Y_i - \bar{Y}$
- The correlation coefficient is a measure of the degree of linear association between the two variables, X and Y
- The value r can assume is from -1 to +1
- When r is +ve, X and Y increase or decrease together,  $r=+1$  means perfect positive correlation, if  $r=-1$ , there exists perfect negative correlation, when  $r=0$ , X and Y are uncorrelated

**RANK CORRELATION COEFFICIENT**

- An assumption made in respect of the linear correlation coefficient is that the variables are quantitative
- If variables cannot be measured numerically, as in ranking effect, we use rank/Spearman's correlation coefficient
- Mechanism is to rank the observation in a sequence eg in order of preference say 1,2 ,3...n
- If 2 variables are so ranked, the rank correlation coefficient can be computed
- $r = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$
- Where D =difference between ranks of corresponding pairs of X and Y, n= number of observations, r ranges between- 1&+1

**CAUTION IN THE USE OF CORRELATION TECHNIQUE**

- The linearity constraint restricts the type of co-variability that can be measured by the technique, since the 2 variables may be strongly connected with a non-linear relationship as in circle
- The technique can not be used to imply any functional relationship between the variables in question
- Correlation technique does not establish or prove any causal relationship between variables
- In correlation analysis, no cause and effect relationship can be inferred
- It seeks to find out if a co-variation exists, but does not suggest that variation in say Y are caused by variation in X or vice versa
- It should be noted that any 2 variables may be correlated for many reasons as follows;
  - X causes change in Y
  - Y causes change in X
  - Neither X nor Y affects each other but both are under the influence of another variable
  - X and Y may be correlated by chance
  - X and Y are jointly dependent (i.e X causes change in Y, and Y causes change in X)

**LINEAR REGRESSION ANALYSIS**

- Regression analysis aims to establish how one variable is related to another, that is, the amount of change in the value of another variable which derives from a unit change in the value of another variable.
- It is based on a causal or functional relationship between variables
- In statistics, regression analysis is a technique which examines the relation of a dependent variable (regressand) to specified independent/explanatory/regressors
- Regression analysis can be used for prediction, inference, hypothesis testing and modeling causal relationships
- The key relationship in a regression is the regression equation
- A regression equation contains regression parameters whose values are estimated using data
- The estimated parameters measure the relationship between the dependent variable and the explanatory variables

#### • **SIMPLE LINEAR REGRESSION**

- A functional relationship is a statement often in the form of an equation of how a dependent variable depends on one or more other variables
- One task in research is to estimate such functional or causal relationship and test the validity or falsity

#### **SIMPLE REGRESSION ANALYSIS**

- Taking the example that consumption of food is a function of family income, implicitly can be stated as
- $C = f(Y)$ , where  $Y = \text{Income}$
- $C$  is the dependent variable, while  $Y$  is the independent variable
- If we assume linear relationship, then  $C = \alpha + \beta Y + e$
- Where  $\alpha = \text{intercept}$  and  $\beta = \text{slope}$  and  $e = \text{random error}$
- Error term is usually posited to be normally distributed
- From sample data  $a$  and  $b$  can be estimated
  
- $\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum xy}{\sum x^2}$
- And  $\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$
- $x$  and  $y$  are deviation of the observations from the means
- Once regression model has been fitted, there is need to confirm the goodness of fit commonly done with  $R$  square and test the statistical significance of the estimated parameters
- Statistical significance of estimated parameters is done using  $t$ -test while the  $F$ -test is used for the over-all fit test
- Prediction can also be done for  $Y$  for given values of  $X$ s

#### **MULTIVARIATE ANALYSIS**

- Real life phenomenon involves interaction between more than 2 variables.
- Where there are more than 2 variables, the analysis is called multivariate analysis.
- There are many multivariate analysis techniques in research to deal with specific phenomenon
- A few commonly used ones will be discussed. They are
- Multiple Regression
- Models of qualitative choice (Probit, Logit and Tobit)
- Analysis of Variance (ANOVA)

#### **MULTIPLE REGRESSION ANALYSIS (MRA)**

- Multiple regression analysis unlike simple regression analysis involves more than one independent variable
- An objective MRA could be to predict the dependent variable or test the relevance of the independent variables in the model/equation
- Note that in the model both dependent and independent variables should be numeric (interval or ratio data)

- However, under special circumstances, the dependent variable and independent variable can use dichotomous variable/dummy variable
- Assuming a linear relationship between the dependent and the independent variables, the model equation can be given as
- $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$
- $\alpha$  and  $\beta$  are parameters of the relationship and the estimates,  $a$  and  $b$  are the standardized partial regression coefficients, interpreted as the change in  $Y$  for a unit change in  $X_1, X_2, \dots, X_k$  holding other variables constant
- Ordinary Least Squares Model, OLS can be used to estimate the parameters
- It minimizes the sums of the squares of the vertical deviations from the line
- Arriving at the estimates involves getting the normal equations, given as

### NORMAL EQUATIONS FOR 2 INDEPENDENT VARIABLES

- $\sum Y = n\alpha + \beta_1 \sum X_1 + \beta_2 \sum X_2$
- $\sum YX_1 = \alpha \sum X_1 + \beta_1 \sum X_1^2 + \beta_2 \sum X_1 X_2$
- $\sum YX_2 = \alpha \sum X_2 + \beta_1 \sum X_1 X_2 + \beta_2 \sum X_2^2$
- There are many software packages available to perform the calculations and output the results.
- Examples are Micro-soft excel, Econometric views, MINITAB etc

### ESTIMATES OF SLOPE PARAMETERS

- Estimates of the slope parameters,  $b_0, b_1$  and  $b_2$  can be got by simultaneously solving the 3 normal equations, as
- $b_0 = Y - b_1 X_1 - b_2 X_2$
- $b_1 = \frac{(\sum X_1 Y_i)(\sum X_2^2) - (\sum X_2 Y_i)(\sum X_1 X_2)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2}$
- $b_2 = \frac{(\sum X_2 Y_i)(\sum X_1^2) - (\sum X_1 Y_i)(\sum X_1 X_2)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2}$
- Note that  $x_i$  and  $y_i$  are deviations from the means
- $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$  which can be used for prediction at given levels of  $X_s$

### FIRST ORDER TESTS

- After producing estimates of the regression model, the next step is to evaluate the model given that a sample data was used in its estimation
- This is done using: F-Statistic,
- T-Statistics
- Coefficient of Multiple Determinations,  $R^2$
- Adjusted R square and
- Standard error
- Coefficient of determination measures the goodness of fit of the model. It measures the explanatory power of the independent variables on the dependent variable
- It is the percentage of the variation in the dependent variable that is explained by the regressors
- Value ranges between 0 and 1
- The higher the value the better the model
- If R-square is zero, then the explanatory variables do not explain any changes in the dependent variable
- The closer R-square is to 1, the better the model estimated
- R-square tends to overestimate the goodness of fit of the model when number of variables and the sample size are increased.
- The adjusted R-square is thus to cater for the over estimation
- It thus gives the useful measure of the goodness of fit
- F-TEST: The F-statistic is used to test whether or not, there is a significant impact between the dependent variable and the independent variables.
- If calculated F is greater than the tabulated F-value, then there is a significant impact between the dependent and the independent variables

EXAMPLE-SEE HAND OUT

• USE OF DUMMY VARIABLES IN MULTIPLE REGRESSION

- A dummy variable is one used as proxy for another variable which cannot be measured quantitatively
- Dummy variable can be used as proxy for qualitative (categorical) explanatory factors such as profession, religion, sex, location, etc
- For example in demand function, we may want to find the effect of say urban and rural residents, then we use dummy say if urban household, then 1, and if rural household, then 0

DUMMY VARIABLES

- Dummy variable is used as proxy to quantitative explanatory variable, when no observations on the variable is available or when it is suitable to do so. For example in the study of estimating savings function and wanting to find the effect of age.
- Age is numeric but can also be stratified in those below 30 years and those above
- The dummy variable will now be those < 30=0, while those>30years=1
- Dummy variable can be used to measure the shift in a function over time, eg Consumption function during war time and normal period.
- It can be used to measure the effect of season, seasonal effect

DUMMY VARIABLES

- Practical example is to consider, consumption pattern in the various quarters of the year
  - Construct the dummy matrix,
- |           | D1 | D2 | D3 |
|-----------|----|----|----|
| • Season1 | 1  | 0  | 0  |
| • Season2 | 0  | 1  | 0  |
| • Season3 | 0  | 0  | 1  |
| • Season4 | 0  | 0  | 0  |
- Note that number of dummy variables used for proxy is always one category less than the total number of categories
  - The category not represented is known as the base category
  - For dummy variables to be meaningful, they must be based on mutually exclusive categories

MODELS OF QUALITATIVE CHOICE

- If independent variable is a qualitative measure rather than a continuous, interval measure, regression analysis would be unreasonable because the estimates will no longer be efficient
- But most of what social scientists study cannot be analyzed with classical regression model, because many attitudes, behaviours, characteristics, decisions and events are measured in discrete, nominal, ordinal variables that are non-continuous
- Analytical technique needed is one that can do the work of multivariate regression analysis without being subjected to the liability of qualitative dependent variable

MODELS OF QUALITATIVE CHOICE

- Several such procedures that can analyze categorical dependent variables exist.
- Examples are Probit, Logit and Tobit models
- They all model the probability of an event, ie how likely an event is to occur
- Qualitative choice analysis methods are used to describe and/or predict discrete categories of variables

PROBIT MODEL

- A way of estimating a dichotomous dependent variable
- Based on the assumption that the variable reflects cumulative normal distribution
- Cumulative Density Function (CDF) of X which follows normal distribution with mean u and variance  $\sigma^2$  is given by
- $f(x) = 1/(\sqrt{2\pi}) * e^{-1/2\{(X-u)/\sigma\}^2}$

## ANALYSIS OF VARIANCE VIS-À-VIS REGRESSION

- ANOVA is a statistical technique developed for the analysis of agricultural experiments by R.A. Fisher
- ANOVA applications assumes a certain experimental design which determines the number of factors, (causes of variation)
- Consider maize yield  $Y$ , produced using fertilizer( $X_1$ ), seed ( $X_2$ ), and type of herbicide( $X_3$ )
- ANOVA breaks down the total variation in maize yield into three separate components; a component due to  $X_1, X_2$  and  $X_3$
- ANOVA is concerned with the variation of  $Y$ , while the values of  $X$ 's are used only for classifying the values of  $Y$  into sub groups
- For example, one group comprising values of  $Y$  corresponding to large values of  $X$ , another group of  $Y$  corresponding to small values of  $X$
  
- The mean value of  $Y$  is estimated for each sub-sample and we have a series of mean values of  $Y$ .
- The logic is that if  $X$ , the basis of classification of  $Y$ 's into sub-samples is an important cause of variation in  $Y$ , there will be large difference between the means of sub-samples
- Such will be indicated by large dispersion of the means of sub-samples ( $Y$ 's) around the pooled mean of  $Y$  (that is large variance of the distribution of the means)
- On the other hand, if  $X$  is not an important source of variation of  $Y$ , there will be small difference between the means of the sub-samples
- In the same vein, this will be indicated by small dispersion of the means of sub-sample( $Y$ 's), around the pooled mean of  $Y$  (that is small variance of the distribution of the means)
- The difference between the means of sub-samples is signified by the value of the variance of the distribution of the means
- There will be 2 estimates of the population variance of  $Y$
- One estimate is obtained by pooling the variances of the subsample and the other from the sampling of the pooled mean • In effect the method of analysis of variance reduces to the estimation of 2 variances, and comparison of these variances to determine if the difference between them is significant
  
- Comparison of the 2 variances is done using F-statistic
- If there are 2 independent variance estimates obtained with  $v_1$  and  $v_2$  degrees of freedom, their ratio has an F-distribution with  $v_1$  and  $v_2$  degrees of freedom
- F-distribution assumes normal distribution of variables
- The greater the difference between the 2 variances, the greater the value of F-distribution and suggests that the difference between the 2 variances is significant or the rejection of the hypothesis that there is no significant difference between the 2 categories

## ANOVA TECHNIQUE ILLUSTRATION

- A researcher wants to find out if there is any significant difference between the means of say 3 populations,  $U_1, U_2, U_3$
- $H_0: U_1=U_2=U_3=U$
- Were the 3 means to be equal, then the 3 populations can be considered as one single population with mean  $U$  and standard deviation  $a$ .  $Y \sim N(U, a)$
- If this is true, the, the sample means  $Y_1, Y_2, Y_3$  should not be significantly different from each other and from the overall mean  $Y$
- If null hypothesis is not true, then the sample means should differ from each other and the pooled mean
- The difference between sample means  $Y_1, Y_2, Y_3$  would be larger than what is due to chance
- Since the estimate of population variance was computed from the difference  $(Y_j - Y)$  it means that the estimate will be large
- This estimate is the cornerstone of the test of difference between means of various samples
- This difference between the sample means is called variation between
- To test the null hypothesis, we compare this estimate with the true population variance

- We reject the null hypothesis, if the divergence between the estimates of the variances is large
- Because we do not know the true population variance, we derive an independent estimate from the sample data as variance =  $\sum (Y_{ji} - \bar{Y}_j)^2 / (N-k)$

The second estimate of population variance is obtained from the sample variances, and this sample variances reflect the variation within each sample

They are unbiased estimates of the population variances

This variance reflects the variation of Y's within the samples, with k-1, degrees of freedom

The ratio of the 2 is an F-distribution

$$F = \frac{\sum nj(Y_j - \bar{Y})^2 / (k-1)}{\sum \sum (Y_{ji} - \bar{Y}_j)^2 / (N-k)}$$

- K = no of sample

$n_j$  = size of jth sample

N = size of the pooled samples

If F calculated > F tabulated, reject null hypothesis

#### GENERAL FORM OF ANOVA

Source of Variation (1)	Sum of Squares (2)	Degree of freedom (3)	Mean Square (4) = 2/3	F (5)	Decision (6)
Between the means	$\sum nj (\bar{y}_j - \bar{y})^2$	$V1 = K-1$	$\sum nj (\bar{y}_j - \bar{y})^2 / (K-1)$		
Within the means	$\sum \sum (y_{ji} - \bar{y}_j)^2$	$V2 = N - K$	$\sum \sum (y_{ji} - \bar{y}_j)^2 / (N-K)$	$\frac{\sum nj (\bar{y}_j - \bar{y})^2 / (K-1)}{\sum \sum (y_{ji} - \bar{y}_j)^2 / (N-K)}$ F1 = -----	
Total Variation	$\sum \sum (y_{ji} - \bar{y})^2$	(N-1)		F from table with $V1 = K-1$ $V2 = N-K$	

- F' = Estimated variance from between the means variation / Estimated variance from within the samples variation
- Both variances are unbiased estimates of the population variance
- Both estimates are independent, hence their ratio has an F-distribution with  $v1 = k-1$ , and  $v2 = N-k$  degrees of freedom
- $N_j$  = size of the jth sample, k = number of samples



ANOVA CONTRASTED WITH REGRESSION

REGRESSION	ANOVA
Similarities	
Total variation in the dependent variable is split into 2 components, the variation explained by the regression line and the unexplained variation, shown by scatter point around the line. In fact $R^2$ indicate the proportion of total variation explained by regression line and its synonymous to additive component principle of ANOVA	It splits the total variation of a variable around its mean into components which may be attributed to specific (additive) causes. Any other variation not so accounted for by the additive causes is assumed to be random or chance variation
Differences	
It examines the variation in the dependent variable Y, as they are affected by variation in the independent variable ,Y	It concentrates on the variation in Y, the values of X are used only to classify T into sub-groups
It provides numerical values for the influence of the various explanatory factors on the dependent variable, in addition to breaking down the total variance of Y into additive components.	It provides only the breakdown of total variance of Y in to additive components, without giving the coefficients indicating influence of Xs on the dependent variable.
COMPLEMENTARITY ANOVA method is used in regression analysis for Test of overall significance of regression . Test of equality of coefficients obtained from different samples.	Test of significance of improvement of fit obtained by the introduction of additional explanatory variables in the function

LIMITATIONS OF ANOVA

- Caution should be exercised in using ANOVA for cross sectional survey data
- ANOVA technique is widely applicable to experimental studies where the number of observations in each category is set equal, which is a balanced design

Chi-Square formula is given by:  $\chi^2 = \sum \{(O-E)^2/E\} = \{(fo-fe)^2/fe\}$

Where O or fo = observed frequency, E or fe = Expected frequency

12. CHI-SQUARE TEST: APPLICATION

AREAS OF APPLICATION OF  $\chi^2$

- Chi –Square test is applicable for these tests
- Test of goodness of fit
- Test of homogeneity of a number of frequency distributions
- Test of significance of association between 2 attributes
- In other words, chi-square test is a test of goodness of fit, homogeneity and independence.
- As a test of goodness of fit, chi square test enables us to see if a distribution of observed data fits the assumed theoretical distribution such as normal distribution, binomial distribution etc
- This is to know if some fitted theoretical distribution fits well the distribution of some observed data

CHI-SQUARE APPLICATION

- Chi-square test can provide answer in this situation,

- If the calculated value of chi-square is **less than** the table value at certain level of significance, the fit is considered to be good, which means that the divergence between the observed and the expected frequencies is attributable to fluctuations of sampling
- But if calculated value is **greater than** its table value, the fit is **NOT** considered to be a good one
- Hypotheses tested are- • H0: There is no difference between fitted distribution and observed data
- H1: There is a difference between fitted distribution and observed data

#### TEST OF INDEPENDENCE

- Chi-square test enables us to explain whether or not 2 attributes are associated.
- We may want to know, whether a new medicine is effective in controlling fever or not
- We proceed on the null hypothesis that the 2 attributes are independent (ie. New medicine and control of fever). Which means that new medicine is not effective in controlling fever.
- On the basis of this, we calculate the expected frequencies and work out the value of Chi-square. If calculated Chi-square is less than table value at a certain level of significance, for a given degree of freedom, we conclude that new medicine is not effective in controlling fever. Our hypothesis thus stands

#### TEST OF INDEPENDENCE

- Chi-square test enables us to explain whether or not 2 attributes are associated.
- We may want to know, whether a new medicine is effective in controlling fever or not
- We proceed on the null hypothesis that the 2 attributes are independent (ie. New medicine and control of fever). Which means that new medicine is not effective in controlling fever.
- On the basis of this, we calculate the expected frequencies and work out the value of Chi-square. If calculated Chi-square is less than table value at a certain level of significance, for a given degree of freedom, we conclude that new medicine is not effective in controlling fever. Our hypothesis thus stands

#### CHI-SQUARE APPLICATION ASSOCIATION

- But if the calculated Chi-square is larger than the table value, we reject the null hypothesis and conclude that, there is association between the 2 attributes.
- The association is NOT because of some chance factor but exists in reality.
- Thus new medicine is effective in controlling fever and thus may be prescribed
- Note however that Chi-square is NOT a measure of the degree of the relationship or the form of the relationship, but simply a technique for judging the significance of such relationship

#### CHI-SQUARE APPLICATION HOMOGENEITY

- Chi-square helps in knowing whether different samples, come from the same universe
- Through this test, we can explain whether the results worked out on the basis of samples are in conformity with well defined hypothesis or the result fails to support the given hypothesis
- The hypothesis of interest thus is-
- H0: The sample is from the population
- Ha: The sample is NOT from the population
- The test thus is an important decision making technique

#### CONDITIONS FOR APPLICATION OF CHI-SQUARE TEST

1. All the members in the sample must be independent
2. No group should contain few members, say less than 10
- In cases where frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups, so that the new frequencies are greater than 10
- Some statistician take this as 5, but 10 is regarded as better
3. The overall number of members/items, i.e N, must be reasonably large, at least 50

#### ILLUSTRATION

- A die was thrown 132 times with these results

- Number turned up 1 2 3 4 5 6
- Frequency 16 20 25 14 29 28
- Test the hypothesis that the die is unbiased
- SOLUTION
- Assuming the die is unbiased, the probability of obtaining any one of the six numbers is 1/6, and the expected frequency of any number coming up is  $1/6 \times 132 = 22$
- We can write the observed frequencies alongside observed and calculate Chi-square

No	Observe	Expect	fo-fe	(fo-fe) <sup>2</sup>	(fofe) <sup>2</sup> /fe
• 1	16	22	-6	36	36/22
• 2	20	22	-2	4	4/22
• 3	25	22	3	9	9/22
• 4	14	22	-8	64	64/22
• 5	29	22	7	49	49/22
• 6	28	22	6	36	36/22

- $\sum \{(fo-fe)^2/fe\} = 9 =$  Calculated value of Chi-square
- Degrees of freedom in the given problem is  $(n-1) = (6-1) = 5$
- Table value of Chi-square for 5 degrees of freedom at 5% level of significance is 11.071.
- Calculated < Tabulated. Accept null hypothesis, thus the die is unbiased. Fluctuation therefore arises from sampling.

#### ILLUSTRATION -2

- Find the value of Chi-square for the following information
- Class A B C D E
- Observe Frequency 8 29 44 15 4
- Expected Frequency 7 24 38 24 7
- Some of the frequencies are less than 10, we need to regroup data and later work out Chi-square
- Class fo fe (fo-fe)<sup>2</sup> (fo-fe)<sup>2</sup>/fe
- A and B  $8+29=37$   $7+24=31$   $6$   $36/31$
- C  $44$   $38$   $6$   $6/38$
- D and E  $15+4=19$   $24+7=31$   $-12$   $144/31$
- Calculated Chi-square = 6.76, At  $df=3-1=2$ , 5% significance level Table Chi-square is 5.991.  $Cal > Table$  thus reject null hypothesis

#### ILLUSTRATION-3

- Genetic theory states that children having one parent of blood types A and the other of blood type B will always be of 3 types, A, AB, B and that the proportions of the 3 types will on the average be 1:2:1
- A report states that out of 300 children having one A parent and one B parent, 30% were found to be the type A, 45% type AB and remaining type B.
- Test the hypothesis that the proposition of genetic theory is supported, ie on the average, type A, AB, B stands in the proportion 1:2:1

#### SOLUTION TO ILLUSTRATION -3

- Observed frequencies of type A, AB, as given in the question are 30% of 300 = 90, 45% of 300 = 135 and 25% of 300 = 75
- The expected frequencies of type A, AB, B as per genetic theory should have been 75, 150, 75 respectively. Tabulating,

Type	Observed Fq	Expected Fq	(fo-fe)	(fo-fe) <sup>2</sup>	(fo-fe) <sup>2</sup> /fe
• A	90	75	15	225	225/75 = 3
• AB	135	150	-15	225	225/150 = 1.5
• B	75	75	0	0	0/75 = 0

- $\chi^2 = \sum \{(fo-fe)^2/fe\} = 3+1.5+0 = 4.5,$

- For  $df=3-1=2$ ,  $\alpha=0.05$ , table  $\chi^2_2 = 5.991$ . Decision is accept  $H_0$

Differences in frequencies are due to chance, and support genetic theory. On the average type A, AB, B prop is 1:2:1

#### ILLUSTRATION-4

- Table shows data obtained during outbreak of smallpox

	Attacked	Not attacked	Total
Vaccinated	31	469	500
Not vaccinated	185	1315	1500
TOTAL	216	1784	2000

Test the effectiveness of the vaccine in controlling small pox, with  $X^2$  at 5% level of significance

#### SOLUTION

$H_0$ : Vaccination is not effective in controlling small pox, ie vaccination and attack are independent

$H_1$ : Vaccination is effective in controlling small pox

#### ILLUSTRATION-4 CONT'D

- Based on the null hypothesis, expectation for a cell (AB) is
- Expectation of cell (AB) =  $(A * B) / N$
- Where A represents Vaccination and B is Small pox attack
- (A) = 500, (B) = 216 and N = 2000
- Expectation of (AB) =  $500 * 216 / 2000 = 54$
- Using the expectation of AB, we can compute table of expectations as:

	Attacked: B	Not Attacked: b	Total
- Vaccinated: A	(AB) = 54	(Ab) = 446	500
- Not vaccinated: a	(aB) = 162	(ab) = 1338	1500
- TOTAL	216	1784	2000

#### CALCULATION OF CHI-SQUARE

Group	ObserveFq	ExpectedFq	fo-fe	(fo-fe) <sup>2</sup>	(fo-fe) <sup>2</sup> /fe
• AB	31	54	-23	529	529/54= 9.796
• Ab	469	446	23	529	529/446= 1.186
• aB	185	162	23	529	529/162= 3.265
• ab	1315	1338	-23	529	529/1338= 0.395

- $\chi^2 = \sum \{(fo-fe)^2 / fe\} = 14.642$
- Degrees of freedom =  $(r-1)(c-1) = (2-1)(2-1) = 1$
- Table value of Chi-square for  $df=1$  at  $\alpha=0.05$  is 3.841
- Calculated value > Table value, thus reject  $H_0$
- We conclude that vaccination is effective in controlling attack of small pox

#### ILLUSTRATION-5

• Two researchers classified some people in income groups on the basis of sampling studies. Their results are as follows;

##### • INCOME GROUPS

Researchers	Poor	Middle	Rich	Total
• A	160	30	10	200
• B	140	120	40	300
• TOTAL	300	150	50	500

• Show that the sampling technique of one of the researchers is defective

##### • SOLUTION

•  $H_0$ : Sampling techniques adopted by the researchers are similar ie no differences between the techniques

If the null hypothesis is true,

Expectation of researcher A clasificatn

- 1.Poor income group =  $200 \times 300 / 500 = 120$
- 2.Middle income group =  $200 \times 150 / 500 = 60$
- 3.Rich income group =  $200 \times 50 / 500 = 20$
- Expectation of researcher B classification of people in
- 1. Poor income group =  $300 \times 300 / 500 = 180$
- 2.Middle income group =  $300 \times 150 / 500 = 90$
- Rich income group =  $300 \times 50 / 500 = 30$

• We can calculate X2 by setting up the table of observed and expected frequencies as follows:

GROUPS	OBSERVED FREQUENCY	EXPECTED FREQUENCY	$(f_o - f_e)$	$(f_o - f_e) / f_e$
Researcher A				
Classifying people as poor	160	120	40	$160/120 = 13.33$
Classifying Middle Income	30	60	-30	$900/60 = 15$
Classifying People as rich	10	20	-10	$100/20 = 5$
Researcher B				
Classifying people as poor	140	180	-40	$1600/180 = 8.88$
Classifying Middle Income	120	90	30	$900/90 = 10$

- $\chi^2 = \sum \{(f_o - f_e)^2 / f_e\} = 55.54$
- Degrees of freedom =  $(c-1)(r-1) = (3-1)(2-1) = 2$
- The table of X2 for 2 degrees of freedom at 5% significant level is 5.991
- The calculated X2 is higher than tabulated value
- We therefore reject the null hypothesis
- This means that the sampling technique adopted by the two researchers differ and thus not similar.
- Therefore technique of one researcher is superior to the other.

This represents the crown of all the efforts that has been put into the research It is an opportunity to put on record your effort at proffering an answer to a problem

### 13. RESEARCH REPORT PREPARATION AND PRESENTATION

#### RESEARCH REPORT PREPARATION

- Research report is the post research description of work done and the results obtained; it is thus an *expost* activity
- It is an over statement to say that this aspect should be done thoroughly.
- Whatever is prepared is the evidence of what has been done
- It is a scientific assessment for which the researcher should take responsibility.
- It is important to ensure the reliability and independence of the analysis that constitutes the report

#### REPORT COMPONENT/SECTIONS

- 1.Preliminary pages
  - a) Cover page; b)Title page; c)Acknowledgements; d)Abstract/Executive summary; e)Table of contents; f) List of tables( where applicable);g) List of figures (where apply)
- 2) Introduction/ Background information

- 3) Literature review
- 4.)Statement of research problem/Research questions
- 5.)Objectives of the research
- 6.)Statement of research hypotheses
- 7.)Significance of research/Research justification
- 8.)Methodology of research ie.a )Research approach/strategy; b) Sampling technique and procedures; c)Data collection- Types of data, sources of data, methods of data collection; d)Analysis of data models and techniques of data analysis and e) Data limitations
- 9.) Results and findings of the research work.-Section should be ordered following the objectives of the work
- 10.) Summary and Conclusions
- 11.) References
- 12.) Appendix, where necessary

#### TENETS AND WRITING TIPS

- Research report could be thesis, dissertations, journal article or conference papers.- Observe the following
- Research purpose should be clear, unequivocal and well understood • Concept should be clarified and placed in proper perspective
- Research design should and procedure should be adequate and appropriate for the objectives. It should be reported in a manner that allows for replication
- Research report should communicate results in objective, measured and transparent manner
- The short coming of the research design and the probable consequences on the finding should be discussed
- The research findings should be based on evidence and analysis from the research
- Researcher should not speculate or report on conclusions not substantiated by the evidence gathered from research
- Writing a report is a cyclical process
- It should start early and continue throughout the work
- Researcher may go forth and back the sections because of the need to make alterations, ensure consistency and logical coherence

#### REASONS FOR REDRAFTING REPORT

- There may be new information that require modification of earlier sections.
- Reasons for redrafting are;-bring in new material, ideas
- Reduce length of paper
- Revise old sections to refer to newly drafted material
- Alter the structure of what has been written
- Respond to suggestions from reviewers
- Remove repetitions

#### WRITING STYLE

- Without good writing style, the content of report will remain poorly understood and thus little policy impact
- Communication is not just about results and findings, but how researcher tells the story of the work.
- It is important to check the logic.
- Promote easier comprehension
- Sentences should be simple, unambiguous and within context
- Some basic principles are;
- Focus on target audience, research community, policy makers, students. Ask questions on whom they are and want
- Be clear as much as possible. Sentences should be short, direct without meandering. Use fewest words possible
- In communicating, the research, simplicity and clarity should be the upper most consideration

- The beginning of report should be provoking enough to grab the interest and curiosity of the audience
- The ending should stimulate action in readership
- Where necessary, report should contain appendices, useful for taking out detailed tables, charts, diagrams, statistical materials which can break up continuity in reading

#### 14. RESEARCH PROPOSAL

Just as research report is ex-post, research proposal is ex-ante. It is a documentation of research plan

##### Research Proposal

- The research plan is the expression of the thinking and direction of the research as conceived by researcher at a point in time.
- Basic questions are
  - What is the nature of the phenomenon which will be investigated?
  - What represents evidence of the entities to be investigated?
  - What topic area is the research concerned with?
  - What is the intellectual puzzle?
  - What is the purpose of the research?

##### • PRELIMINARY STEPS TO TAKE

- Choose an area of research
- Read relevant literature and discuss with other researchers
- Choose the topic and title of the research

##### COMPONENTS OF RESEARCH PROPOSAL

- The components of a research proposal address these research issues
  - What is the problem to be investigated?
  - Why should the problem be studied?
  - What will be achieved through the research?
  - What information is needed to meet the research objectives?
  - How will the research process be carried out?
  - What type of results are expected from the research and who are the potential beneficiaries?
  - How will the utilisation of results be ensured?
  - Who will do what and when will it be done?
  - What resources are needed to carry out the research?

##### SECTIONS OF PROPOSAL

- Summary/Abstract-It's an encapsulation of what is contained in the proposal • Even though it comes first in proposal, it is usually written last.
- Its recommended that it should not be more than 300- 500 words
- Should be written in a manner to excite, enthuse, interest or convince reader of the significance of the research
- Should contain info on problem,objectives, procedures, methods used,resource needs and likely outcomes and benefits of the research

##### Table of contents

- It is the listing of major contents of the research proposal
- It outlines the sections, sub sections and give the pages where they are found
- 3.INTRODUCTION/BACKGROUND INFORMATIO
- Its devoted to stating research significance
- Should be short and captivating and inform the readers of the reasons for doing the research

### 3. PROBLEM STATEMENT

- It represents the reasons behind your proposal and specifies what is intended to be changed/improved through research
- It should address these questions
- What is the need addressed by the researcher? ie gaps research address?
- Why is it important to address the gaps?
- What is known about the problem?
- How will the research impact the problem?
- What specific questions does the work seek to provide answers to

### EXAMPLES OF RESEARCH PROBLEM

- Access to capital is important to reducing shortages of capital among micro entrepreneurs. Knowledge of their socioeconomic and institutional factors that influence access to formal credit is important for improving the effectiveness of credit programs. However, there is limited analysis of these factors and their interacting effects on credit access. This gap in knowledge is a research problem which poses knowledge need

#### 3. Statement of research Questions

- Research questions pose knowledge needs and intellectual puzzle justifying the research
- Research questions are short and precise statements of knowledge requirements to be answered by the investigations.
- Examples are questions on relationships among variables

### 5. WORKING HYPOTHESES

- They are tentative assumptions made in order to draw out and test empirical manifestations
- It delineates and focuses the research
- It shapes the type of data to collect, the analysis to do and it affects the manner in which empirical tests will be conducted
- There is a null and alternative hypothesis

#### 4. LITERATURE REVIEW

- Main function is to;
- Give reasons why the topic is sufficient importance
- Provide researcher with up to date account and discussion of literature on the issues relevant to the topic.
- Provide conceptual and theoretical context in which topic can be situated
- Discuss relevant research carried out on the same topic

### 7. JUSTIFICATION OF STUDY

- Section identifies the expected impact of the research
- It indicates outcomes in terms of contribution to- Solution of specific problem on which research is focused
- The process of policy formulation/implementation
- Empirical research in the field of enquiry
- Developmental process at the local/state/national

### 8. OBJECTIVES OF RESEARCH

- Objectives are measurable outcomes of research, tangible results of research activities
- They are what the research will accomplish
- It is the benchmark against which the work will be judged as successful
- Could be broad or specific
- Research objective reflect the purpose of the work which could be exploratory, descriptive, diagnostic, experimental or a mixture
- It links theoretical relationship implied by hypothesis to analytical procedures and provide orientation needed to carry out the research



#### 9. RESEARCH DESIGN

- Section contains information on research design, sampling, measurement, data collection/management and data analysis
- Purpose of the section is to describe how the research will be carried out .INVOLVES THESE
- Precise and concise statement of the research approach to be used
- Description of elements, variables to measure and how it will be done
- Description of the population, samples to use, sampling selection procedures, sampling units and sample size
- Description of the approaches and methods to be applied in collecting primary and secondary information, indicating procedures and instruments to be used and sources of information.
- Description of types of analysis to be carried out, the procedures for processing and analyzing information and type of facilities to use
- Statement of hypotheses-testing modalities

#### 10. BUDGET

- This is the statement of proposed expenditures
- Involves costing the activities of the research
- It sets out the resource implications and funding needs of work
- Includes direct and indirect expenses
- Direct are-travel costs, documentation, materials, equipments, professional and support services costs
- Indirect are institutions overhead costs eg. rent, utilities, salaries and wages of staff
- Necessary to be realistic and credible

#### 11. WORK PLAN

- Funding agencies require work plan and time table
- Here work plan sets out activities, deliverables and milestones
- Time table is schedule of work according to time of implementation
- Can use Ghant chart to indicate

#### 12. REFERENCES AND APPENDIX

- References should be used to justify and support arguments, make comparisons between different research works
- APPENDIX-supply additional information that is not directly part of the research proposal

#### PRINCIPLES IN WRITING RESEARCH PROPOSAL

- Proposal must address a research problem that is significant
- It should be internally consistent (synergy between hypothesis, objectives, and methodology)
- Proposal must include procedures for addressing every objective, hypothesis or research questions
- Must contain a clear description of theoretical base of methodology to be used
- Proposal should be presented in a manner that flows logically from section to section

#### PRINCIPLES IN WRITING RESEARCH REPORT

- Proposal should be in a language and editorial style appropriate to the scientific discipline
- Proposal should be easily readable with clear and concise language
- It is a documentation of scholarship, innovativeness and effort of the researcher
- Proposal should be self-contained and self recommending with sufficient details to convey intent
- The whole process is interactive, continuously you go forth and back.