SLIIT

COMPUTING | BUSINESS | ENGINEERING

Research Methods

Lecture 1-1: Introduction and Types of Research

Unit Learning Outcomes

By the end of the module you will be able to:

1. Identify and articulate a research question.
2. Critically evaluate literature particularly with regard to its contribution to research.
3. Appreciate and be able to evaluate a range of research methodologies, tools and techniques that can be applied to a particular area of research.
4. Reflect on the research process and its outcome.

How do we define research?

- Research may be defined as a careful and systematic process of inquiry to find answers to problems of interest.

- To do "research" is to investigate a problem systematically, carefully, and thoroughly.

- This requires the researcher to follow a process, a sequence of steps from formulating the research problem to publishing the results.

- This process is called the scientific method, or more appropriately, "scientific methods" because there is more than one way of doing this.

What is research ?

- Careful study and investigation, especially in order to discover new facts or information" (1)
- A voyage of discovery; A journey; An attitude; An experience; A method of critical thinking; A careful critical enquiry in seeking facts for principles. (2)
- Scientific research differs from other kinds of research in that it is a continued search for scientific knowledge and understanding by scientific methods. (3)

(1) Oxford paperback dictionary, 3rd edition
(2) Research Methodology by Dr. M S Sridhar, Head, Library & Documentation, ISRO Satellite Centre, Bangalore 560017
(3) THE PRINCIPLES OF SCIENTIFIC RESEARCH, By PAUL FREEDMAN, B.Sc., M.I.E.E., F.I.E.S., Head of Lamp Research, Messrs. Crompton Parkinson, Ltd.

What is research?

Uma Sekaran **defines research as**

"an *organized, systematic, data-based, critical, objective, scientific inquiry or investigation* into a specific problem, undertaken with the purpose of finding answers or solutions to it. "

Research provides the needed information that guides you to make *informed* decisions to successfully deal with problems.

RM

Theory of how RESEARCH should be undertaken

Goals of research

- **Establishing facts**
- **Analyzing information**
- **Reaching new conclusions**

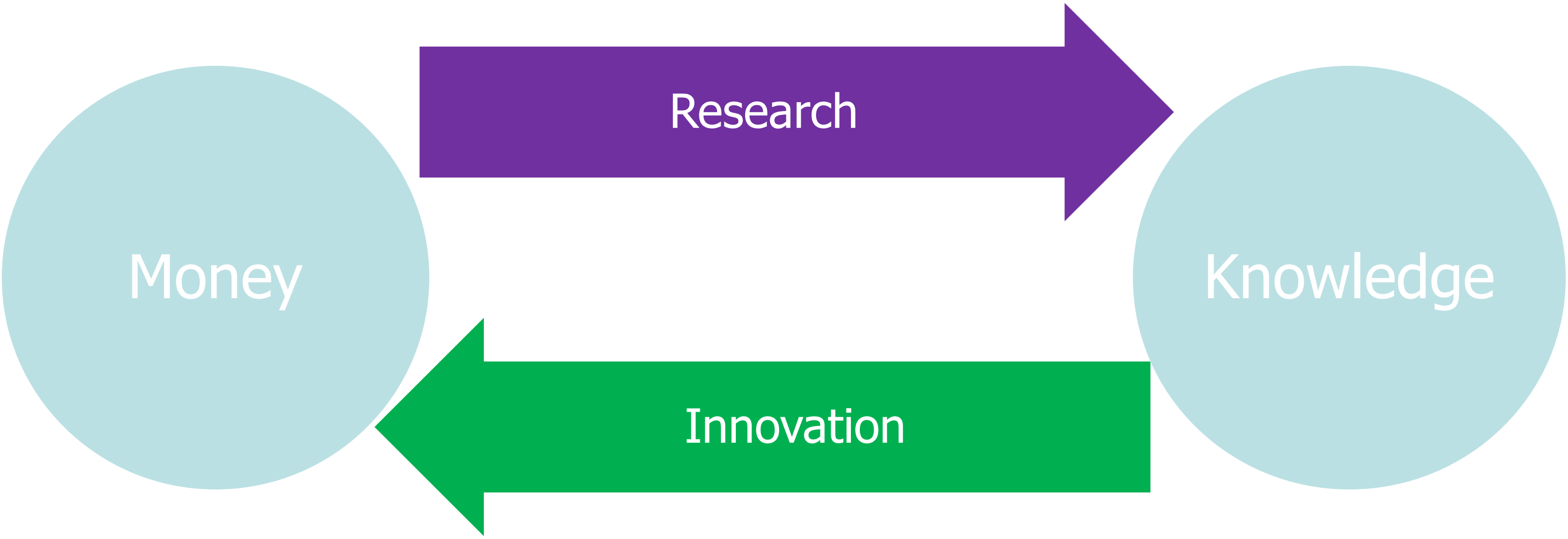
Useful terms and sample findings

- **Identify: Problems, opportunities, criteria**
- **Define: Concepts, requirements**
- **Describe: Process, usage, environments**
- **Explore: Perceptions, reactions**
- **Generate: Ideas, hypothesis, alternatives, explanations**
- **Evaluate: Feasibility, attractiveness, support**
- **Select: Theme, service, product, concept, advert**
- **Test: Assumptions, preferences**
- **Measure: Size, growth, frequency**

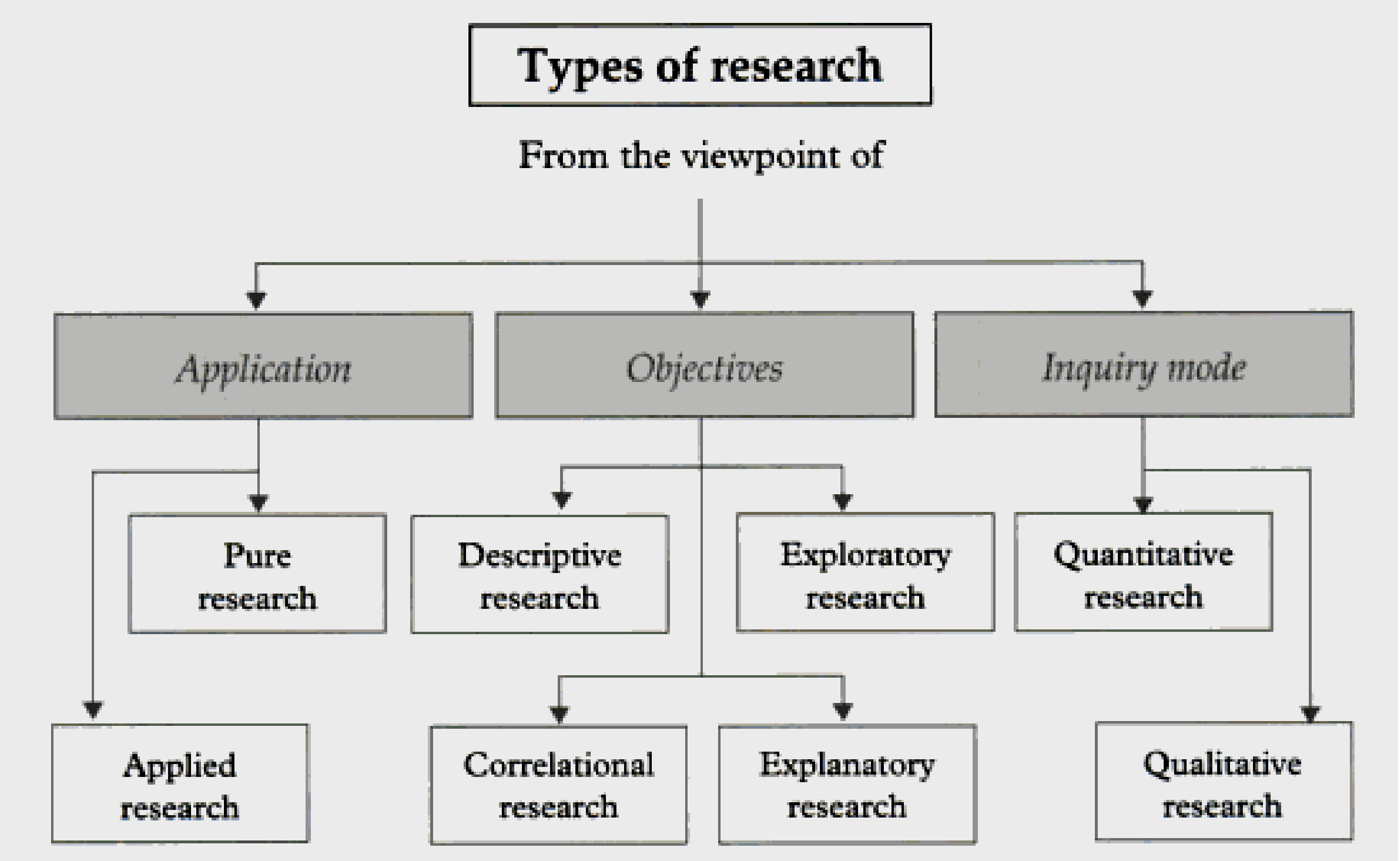
Innovations

- **Definition:**

  The process of translating an idea or invention (research) into a good or service that creates value or for which customers will pay.



Types of research



Types of Research: Application

Pure Research Vs Applied Research

Pure Research:

Primary objective is the *advancement of knowledge* and the theoretical understanding of the relations among variables.

It is exploratory and often driven by the researcher's curiosity, interest, and intuition

Applied Research:

That is research accessing and using some part of the research communities' accumulated theories, knowledge, methods, and techniques, for a specific, often commercial, or client driven purpose (good/process)

Pure Research Vs Applied Research

Pure research is for the sake of *curiosity and functions to advance knowledge* for its own sake. This type of research is usually carried out in government-funded projects by University research facilities or specific government laboratories.

Applied research is for the sake of *technological advancements*. This research anticipates that the results found will lead to the development of commercially viable goods or processes.

There is a **relationship** between these two types of research: pure research *generates new ideas* and applied research *takes these ideas to create new inventions*. Then, through development, the new inventions are transformed into commercial products. This is a simple model to understand the different types of research and the eventual development of new products

Types of Research: Objectives

Descriptive:

Attempts to describe systematically a situation, problem, or phenomenon

Correlational:

A study is to discover or establish the existence of a relationship/association/interdependence between two or more aspects of a situation or phenomenon

Explanatory:

Attempts to clarify why and how there is a relationship between two aspects of a situation or phenomenon

Exploratory:

A study is undertaken with the objective of exploring an area where little is known

Types of Research: Objectives

Descriptive:

This describes phenomena as they exist. It is used to *identify and obtain information on the characteristics of a particular issue*. It may answer such questions such as:

What is the absentee rate amongst a particular group of workers?
What are the feelings of workers faced with redundancy?

The data collected are often quantitative, and statistical techniques are usually used to summarise the information. Descriptive research goes further than exploratory research in examining a problem since it is undertaken to ascertain and describe the characteristics of the issue.

Types of Research: Objectives

Explanatory:

This is a continuation of descriptive research. The researcher goes beyond merely describing the characteristics, to *analyse and explain why or how something is happening*.

Thus, explanatory (analytical) research aims to understand phenomena by *discovering and measuring causal relations* among them.

Root cause analysis, co-relational analysis, hypothesis testing are very much useful

Types of Research: Objectives

Exploratory:

This is conducted when there are *few or no earlier studies* to which references can be made for information.

The aim is to look for patterns, ideas or hypotheses rather than testing or confirming a hypothesis.

In exploratory research the focus is on gaining insights and familiarity with the subject area for more rigorous investigation later.

Types of Research: Inquiry Mode

Quantitative Research

Quantitative research tries to *quantify the variation* in a phenomenon, situation, problem or issue

Information is gathered using predominantly *quantitative variables, usually in interval or ratio scale* and analysis is geared to ascertain the *magnitude of the variation*

Qualitative Research

Purpose of the study is primarily to *describe* a situation, phenomenon, problem, or event

Information is gathered through the use of variable measured on *nominal* or *ordinal* scales (*qualitative measurement scales*)

Types of Research: Inquiry Mode

Quantitative Research -*structured*

Research process is *predetermined*.

More appropriate to determine the *extent*

They will favour methods such as *surveys and experiments*, and will *attempt to test hypotheses* or statements with a view of generalising from the particular. This approach typically concentrates on *measuring or counting and involves collecting and analysing numerical data and applying statistical tests*.

Classification of quantitative research

- Inferential

  Inferential means survey research where a sample of population is studied (questioned or observed) to determine its characteristics, and it is then inferred that the population has the same characteristics

- Experimental

  Experimental approach is characterized *by* some variables are manipulated to observe their effect on other variables

- Simulation approaches to research

  Simulation approach involves the construction of an artificial environment within which relevant information and data can be generated

Types of Research: Inquiry Mode

Qualitative Research-*unstructured*

The alternative tradition is the qualitative approach.

Here the investigator views the phenomena to be investigated as more personal and softer.

Research process is more flexible.

More appropriate to explore the *nature*

Table 1.1 Types of research studies from the viewpoint of objectives

Examples

- Socioeconomic characteristics of residents of a community
- Attitudes of students towards quality of teaching
- Types of service provided by an agency
- Needs of a community
- Sale of a product
- Attitudes of nurses towards death and dying
- Attitudes of workers towards management
- Number of people living in a community
- Problems faced by new immigrants
- Extent of occupational mobility among immigrants
- Consumers' likes and dislikes with regard to a product
- Effects of living in a house with domestic violence
- Strategies put in place by a company to increase productivity of workers

---

- Impact of a program
- Relationship between stressful living and incidence of heart attacks
- Impact of technology on employment
- Impact of maternal and child health services on infant mortality
- Effectiveness of a marriage counselling service on extent of marital problems
- Impact of an advertising campaign on sale of a product
- Impact of incentives on productivity of workers
- Effectiveness of an immunisation program in controlling infectious disease

---

- Why does stressful living result in heart attacks?
- How does technology create unemployment/employment?
- How do maternal and child health services affect infant mortality?
- Why do some people have a positive attitude towards an issue while others do not?
- Why does a particular intervention work for some people and not for others?
- Why do some people use a product while others do not?
- Why do some people migrate to another country while others do not?
- Why do some people adopt a program while others do not?

Go to next pg

Aim	*Main theme*	*Type of research*
To describe what is prevalent regarding: • a group of people • a community • a phenomenon • a situation • a program • an outcome	**To describe what is prevalent**	Descriptive research
To establish or explore: • a relationship • an association • an interdependence	**To ascertain if there is a relationship**	Correlational research
To explain: • *why* a relationship, association or interdependence exists • *why* a particular event occurs	**To explain why the relationship is formed**	Explanatory research

Definitions of originality

- **Saying something nobody has said before**
- **Carrying out empirical work that has not been done before**
- **Synthesizing something that has not been put together before**
- **Making a new interpretation of someone else's material or ideas**
- **Taking a new technique and applying it to an existing area**
- **Taking an existing technique and applying it to a new area**
- **Continuing a previously original piece of work**
- **Being cross-disciplinary and using different methodologies**
- **Testing existing knowledge in an original way**

The characteristics of Scientific Research

1. **Purposiveness**
2. **Rigour**
3. **Testability**
4. **Replicability**
5. **Precision and Confidence**
6. **Objectivity**
7. **Generalizability**
8. **Parsimony**

Eg: Consider the case of a manager who is interested in investigating how employees commitment to the organization can be increased.

How do the above hallmarks apply to this research so that it could be considered scientific

Purposiveness

Research has to be started with definite aim or purpose.

Eg:

A manager focuses on increasing the commitment of employees to the organization.

An increase in employee commitment will translate into less labour turnover, less absenteeism and probably increased performance levels, all of which will definitely benefit the organization.

Therefore, the research has a purposive focus.

Rigour (strict precision)

A theoretical base and a sound methodological design would add rigour to a purposive study.

Rigor connotes carefulness, scrupulousness, and the degree of exactitude in research investigation

If the manager asks only some of the employees to indicate what would increase their level of commitment, and reaches several conclusions, is the research rigorous enough to be scientific?

Testability

Hypothesis could be tested by applying certain statistical tests to the data collected for the purpose.

If after talking to a random sample of employees and study of the previous research done in the area of organizational commitment, the manager or researcher develops a certain hypothesis on how employee commitment could be enhanced, then these can be tested using certain statistical tests to the data collected for this purpose.

Replicability

The results of the tests of hypotheses should be supported again and yet again when the same type of research is repeated

Eg:

If a certain finding from a research in a particular organization is similar to the findings from another organization after conducting the same research credibility and the confidence of the scientific nature of the research is enhanced.

Precision and Confidence

Precision refers to the closeness of the findings to reality.
Confidence Interval in statistics is referred here as precision.

Confidence refers to the probability that our estimates are correct. Example: 95% of the time our results would be true. This is known as Confidence Level.

Objectivity

The conclusions drawn through the interpretation of the results of data analysis should be objective;

that is,

they should be based on the facts of the findings derived from actual data, and not on our own subjective or emotional values.

Generalizability

Generalizability refers to the scope of applicability of the research findings in one organizational setting to other settings.

The more generalizable the research, the greater its usefulness and value.

Eg:

If the findings that "participation in decision making enhances organizational commitment" are found to be true in a variety of settings such as in manufacturing, industrial, service sector etc., then the generalizability of the findings is enhanced.

Parsimony

Simplicity in explaining a phenomena or a problem and in generating solutions for the problems, is always preferred to complex research frameworks that consider an unmanageable number of factors.

Unit Administration

Plagiarism

- Plagiarism means presenting work done (in whole or in part) by someone else as if it were one's own
- Is an academic dishonesty

If anybody found guilty

- Maximum action would be taken against the student/s under the SLIIT rules and regulations

And if many involved, all are equally responsible

- Impossible to adjudge who did?
- Assessment marks would be simply 0
- No repeat attempt is allowed

THANK YOU

Research Methods

Lecture 1-2: The Research Process

Agenda

- Building Blocks of Science in Research
- The Research Process

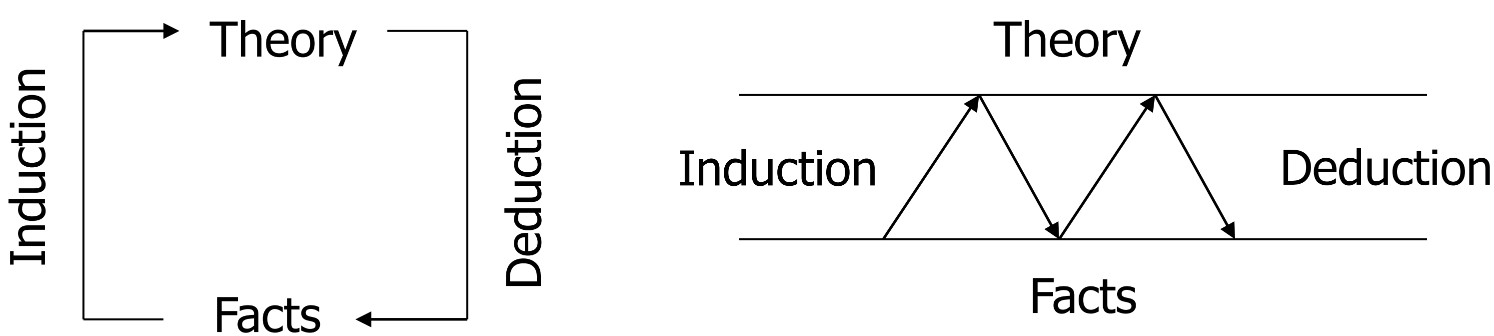
Building Blocks of Science in Research

Deduction and Induction

Induction is the movement or generalization from facts ("observations", "empirical world" or "reality") to theory

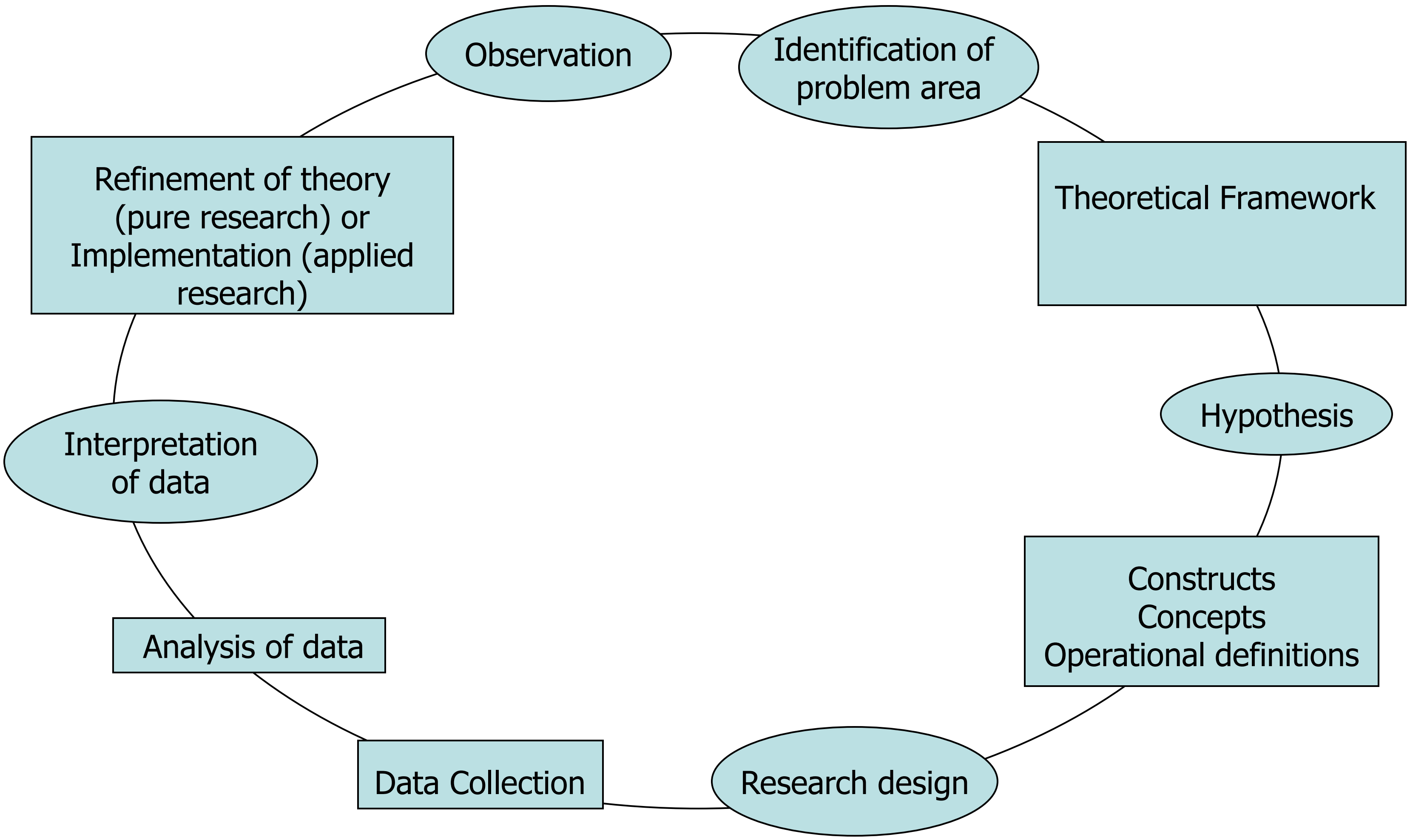
While,

deduction is the reverse appeal to facts to test theories. Thus empirical regularities suggest theories and theories are in turn tested against facts.



The interaction between facts and theory

Building Blocks of Scientific Inquiry



Example

A sales manager might *observe* that customers are perhaps not so pleased as they used to be.

This process of observation or sensing the phenomena around us, is what gets most research (applied or basic) started.

The next step is to determine whether there is a real problem, and if so, how serious it is.

This problem identification calls for some preliminary data gathering.

Example (cont…)

The manager might talk casually to a few customers to find out how they feel about the products and customer service.

Combining information from formal and informal interviewing has helped the manager to determine that a problem does exist.

It helps the manager to formulate a conceptual model or theoretical framework of all factors contributing to the problem.

Example (cont…)

From the theoretical framework, which is a meaningful integration of all the information gathered, several hypotheses can be generated and tested to determine if the data support them

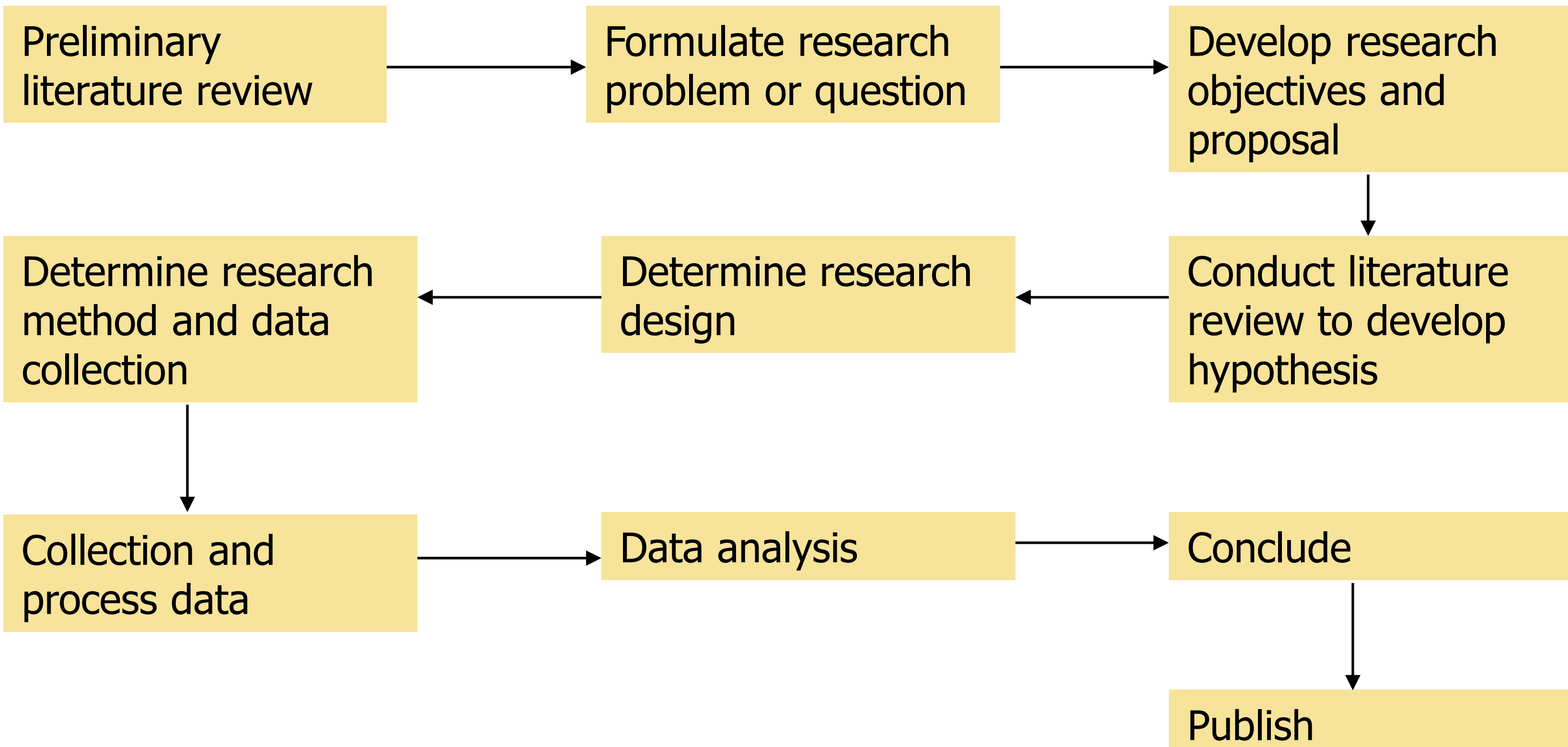
Concepts are then operationally defined so that they can be measured.

A research design is set up to decide among other issues how to collect further data, analyze and interpret them and finally, to provide an answer to the problem.

The Research Process

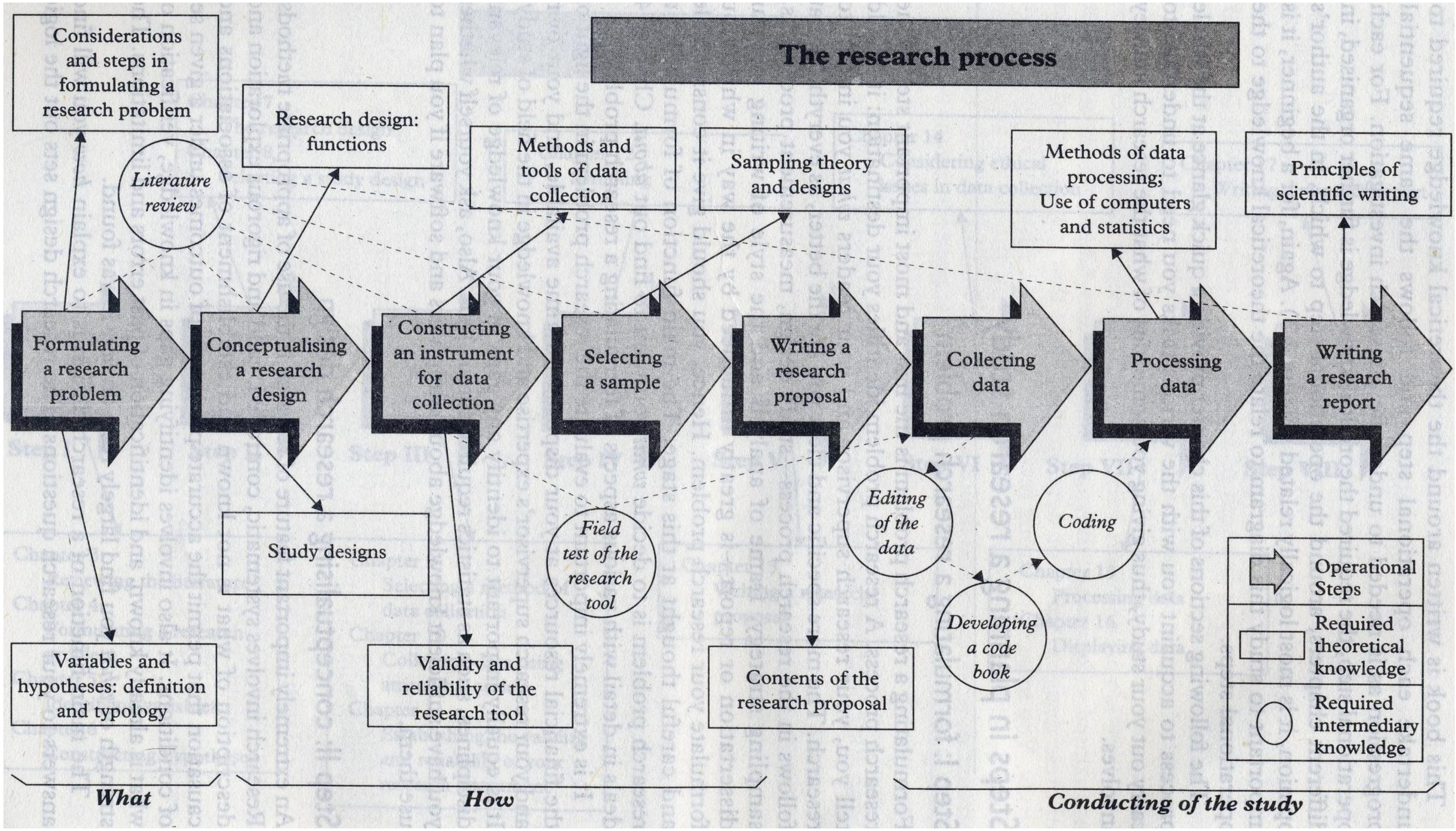
Research has to be systematic and it should follow a series of steps called the research process.

The research process



Source: Practical Research Methods by Willie Tan, 2nd Edition, Prentice Hall, 2004

The Research Process



The Research Process

Formulating a Research Problem

- Identifies researchers' destination
- Explains "What researcher intends to research."

Conceptualizing a research design

- Involves identifying gaps in knowledge, verification of what is already known, and identification of past errors and limitations
- Include the following: Study design, the logistical arrangements that you propose to undertake, the measurement procedures, the sampling strategy, the frame of analysis and the time-frame

The Research Process

Constructing an Instrument for Data Collection

- Anything that becomes a means of collecting information for your study is called 'research tool' or a 'research instrument'
- Construction of a research tool is the first step in carrying out a study
- Planning to collect data (primary data) or use existing data (secondary data)

Selecting Sample

- Selection of relatively small number of units which provides a fairly true reflection of the sampling population with a sufficiently high degree of probability

The Research Process

Writing a Research Proposal

- Lay everything together in a way that provides adequate information, for your research supervisor and others, about the research study

-Is a overall plan which describes research problem, plans and methods of investigation

-Details of the operational plan for obtaining answers to your research question

-It should explain

-What you are proposing to do?

-How you plan to proceed?

-Why you selected the proposed strategy?

The Research Process

Collecting Data

Processing Data

- -Method of analysis depends on
  - -Type of Information: descriptive, quantitative, qualitative or attitudinal
  - -Method of communicating findings to the readers

Publication

- -Should be written in academic style

Other types of research

Case Studies

Case studies involve, in depth, contextual analysis of similar situations in other organizations, where the nature and definition of the problem happen to be the same as experienced in the current situation.

Hypotheses can be developed in case studies as well. However, if a particular hypothesis has not been substantiated in even a single other study, no support can be established for the alternate hypothesis developed.

Other types of research

Action Research

Action research is sometimes undertaken by consultants want to initiate change processes in organizations.

In other words, action research methodology is most appropriate while effecting the planned changes.

Here, the researcher begins with a problem that is already identified, and gathers relevant data to provide a tentative problem solution.

The solution is then implemented, with the knowledge that there may be unintended consequences.

The effects are then evaluated, defined and diagnosed.

THANK YOU

Research Methods

Lecture 1-3: Research proposals

Types of proposals

- **Research proposals**
  - Goals/objectives and scope
  - Equipment requirement

- **Project proposals**
  - Goals/objectives
  - KPI and its measurements
  - Novelty/smart or innovative in nature

Before writing a project proposal

- **Critical Thinking**
  - Objective analysis and evaluation of an issue in order to form a judgment.
- **Identify a research problem**
  - Relevant to the study
  - Interesting to the student
  - Have an adequate depth of research/quantum of work
- **Limit scope/demarcate boundaries**
  - Objectives should be properly set
- **Qualify the idea to a project proposal**

Structure for brief/simple proposal

- **Cover page (format)**
- **Project proposal summary sheet (in given format)**
- **Introduction**
  - Summary of the problem
  - Briefly mention the problems and importance
- **Goals and objectives**
  - Objectives should be smart
- **Project design**
  - Composed of multiple elements and describe later
- **Conclusion**
  - Summarize: problem, motivation, proposed solution, planned important contribution
- **References**

Structure for a detail proposal

- **Title page**
- **Declaration form**
- **Abstract**
- **Table of contents**
- **Background of research**
- **Literature review (theoretical background, empirical studies)**
- **Research problems**
- **Research question/hypothesis**
- **Goals and objectives**
- **Importance of research**
- **Scope of the research**

- **Conceptual framework**
- **Mathematical or other model if relevant**
- **Operationalization**
- **Research methodology**
  - population and sample including sample size calculation and sampling techniques, acceptance and reject criteria in sampling,
  - types of data and data collection methods,
  - pilot study if applicable,
  - reliability and validity,
  - data presentation and analysis methods .tools etc.
- **Limitations of the research**
- **Structure of the dissertation,**
- **References**
- **Time frame**
- **Budget**

Crafting goals and objectives

- **Goal is broad in scope and also in generic form while objectives are more specific**
- **One goal and multiple objectives (recommended 3 to 5) which should be SMART**
  - ***S****pecific*
  - ***M****easurable*
  - ***A****ttainable/* ***A****chievable*
  - ***R****ealistic*
  - ***T****imely/* ***T****ime bound*

Time line

- **Gantt chart is a good option**
  - *Duration of events are denoted by the starting and ending points of bars*
  - *Events can be parallel*
  - *Gantt chart software (Microsoft project, Gantt Project) or excel add in can be used*
  - *Events should be clearly demarcated*
  - *In progress review, two colours (for instance green and red) can be used in bars to indicate the level of completion*
  - *Horizontal axis will be the time*

Citing and referencing

- **Citing**
  - *refer to (a passage, book, or author) as evidence for or justification of an argument or statement, especially in a scholarly work*
  - *Citation style (eg:- IEEE, Harvard, APA)*
- **Referencing**
  - *List of references in alphabetical order or the order of citation*
  - *Referencing style (eg:- IEEE, Harvard, APA)*
  - *Use of reference management tool/software (eg: -Mendeley, Zotero, refer wikipedia for more with comparison)*
  - *Should adequately include recent references*
  - *Majority should not be internet references*
  - *Conference papers, journal papers, books, magazines, manufacturer's handbooks/manuals/newsletters, annual statements, patents in case products, technical notes, application notes, data sheets, e-resources*

What is plagiarism

- **Plagiarism is one specific form of cheating. In simple terms it is a claim somebody else's work, results as yours without giving due recognition to the owner.**
- **An interesting to note that one dictionary defines a plagiarist as a kind of thief: “one who steals the thoughts or writings of others and gives them out as [his/her] own”.**
- **In order not to plagiarize it is essential to refer to the work of others whenever you are influenced by, build on, or make claims based on their work.**

Tips for not to plagiarize

- **Citing the reference correctly**
- **Writing a critical review**
- **Using quotations**
- **Using paraphrasing**
  - 'A rewording of something written or spoken by someone else, esp. with the aim of making the sense clearer; a free rendering of a passage' (Oxford English Dictionary 2007).
- **Referring to common knowledge**
  - Common knowledge is what "everybody knows"(Wikipedia 2007)
  - 'Common knowledge: facts that can be found in numerous places and are likely to be known by a lot of people' (Anon 2004).

THANK YOU

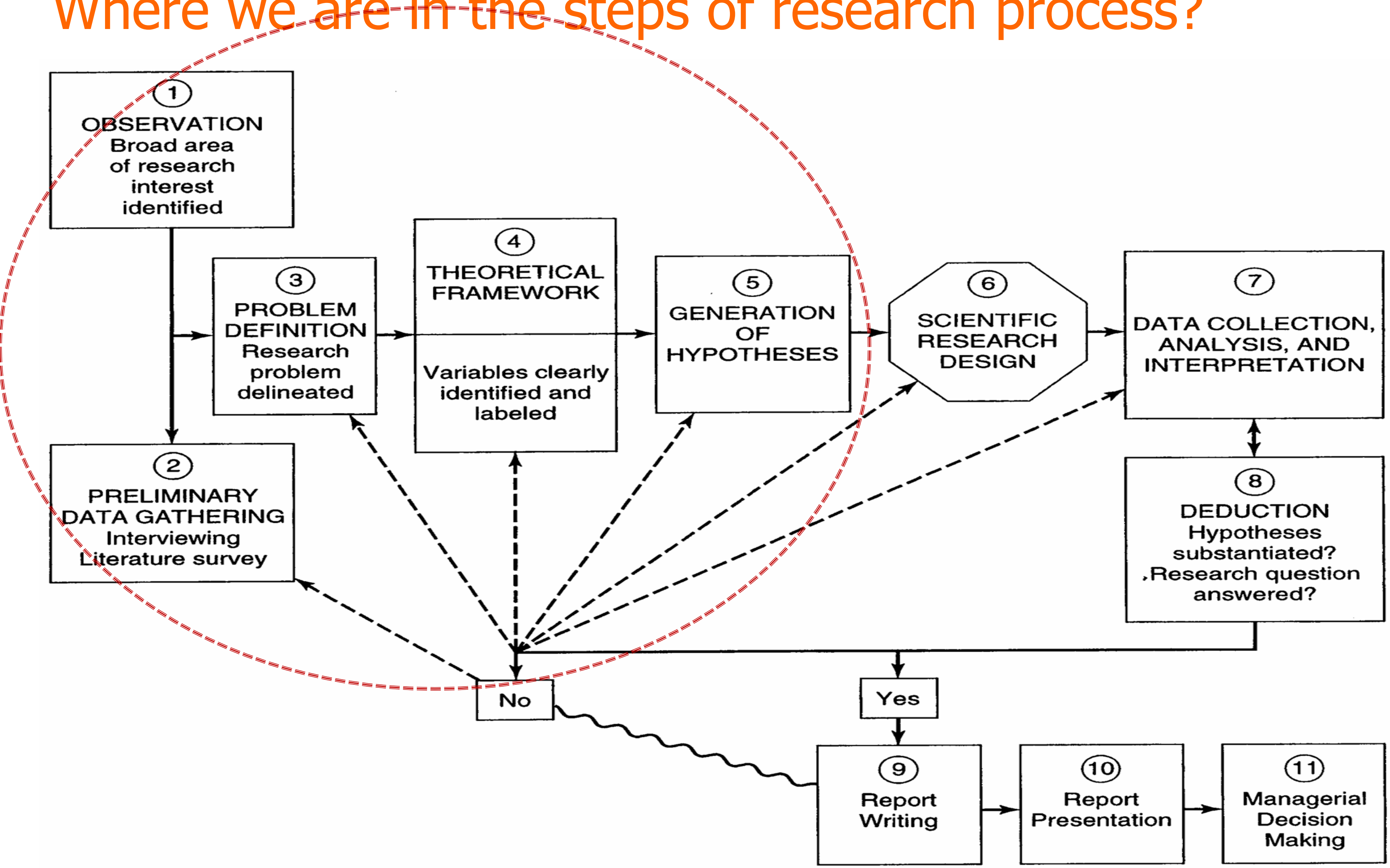
Research Methods

Lecture 2-1: Literature Survey, Theoretical Framework and Generation of Hypothesis

Agenda

- Literature Survey
- Problem Definition
- Research Objectives
- Theoretical Framework
- Hypothesis generation and testing

Where we are in the steps of research process?



Literature Survey

Research has to be done to contribute new knowledge to the existing body of knowledge

New knowledge is:

A solution to an existing problem (complains regarding the quality)

An area where improvement is required (Improve accuracy of character recognition)

Understanding certain phenomena

An answer reached by empirical study

Literature survey is an analysis of existing body of knowledge in your area of interest

Literature Survey

Literature survey is the documentation of a comprehensive review of the published and unpublished work from secondary sources of data in the areas of specific interest to the researcher

Primary Data: Data gathered for research by your own through observations, interviews, questionnaires

Secondary Data: Data gathered through existing sources such as statistical bulletins, government publications, data available from previous research work (Evaluation criteria: timeliness, accuracy, relevancy, cost)

The higher the academic level of your research, the more important a thorough integration of your findings with existing literature

Literature Survey

Literature survey helps you to:

-Bring clarity and focus to your research problem (RP)

idea of the RP -> LS -> Shaping your RP

Relationship between RP & body of knowledge in the area

-Improve your methodology

Methodologies that have been used by others

Can avoid problems and pitfalls others faced

-Broaden your knowledge base in your research area

Theories put forwarded and existing gaps can be identified

-Contextualize your findings

what others have found? Difference between your one and others? Your contribution to the existing body of knowledge?

Features of literature review

- A literature review is a systematic way of showing evidence of your reading and how it relates to your investigation

- Literature review needs to be
  - essentially a critical analysis (be logical and intelligent)
  - focus towards the identification/derivation of research gap
  - cover the breadth adequately
  - include most recent literature

- Structuring can be
  - chronological at beginning (follows historical development)
  - thematic (categories or contexts)
  - classic studies (major/key writings about area)
  - inverted pyramid (starting from broad aspect and narrow down to the focused)

Critical analysis

- What we expect from you:

  We want you not just to find and know information, but we need to see how you analyse and evaluate the information and ideas.

  We want you to come to some conclusions about it, so that we get your views and opinions.

  And these must be supported by evidence from the reading you have done.

Ways of being critical

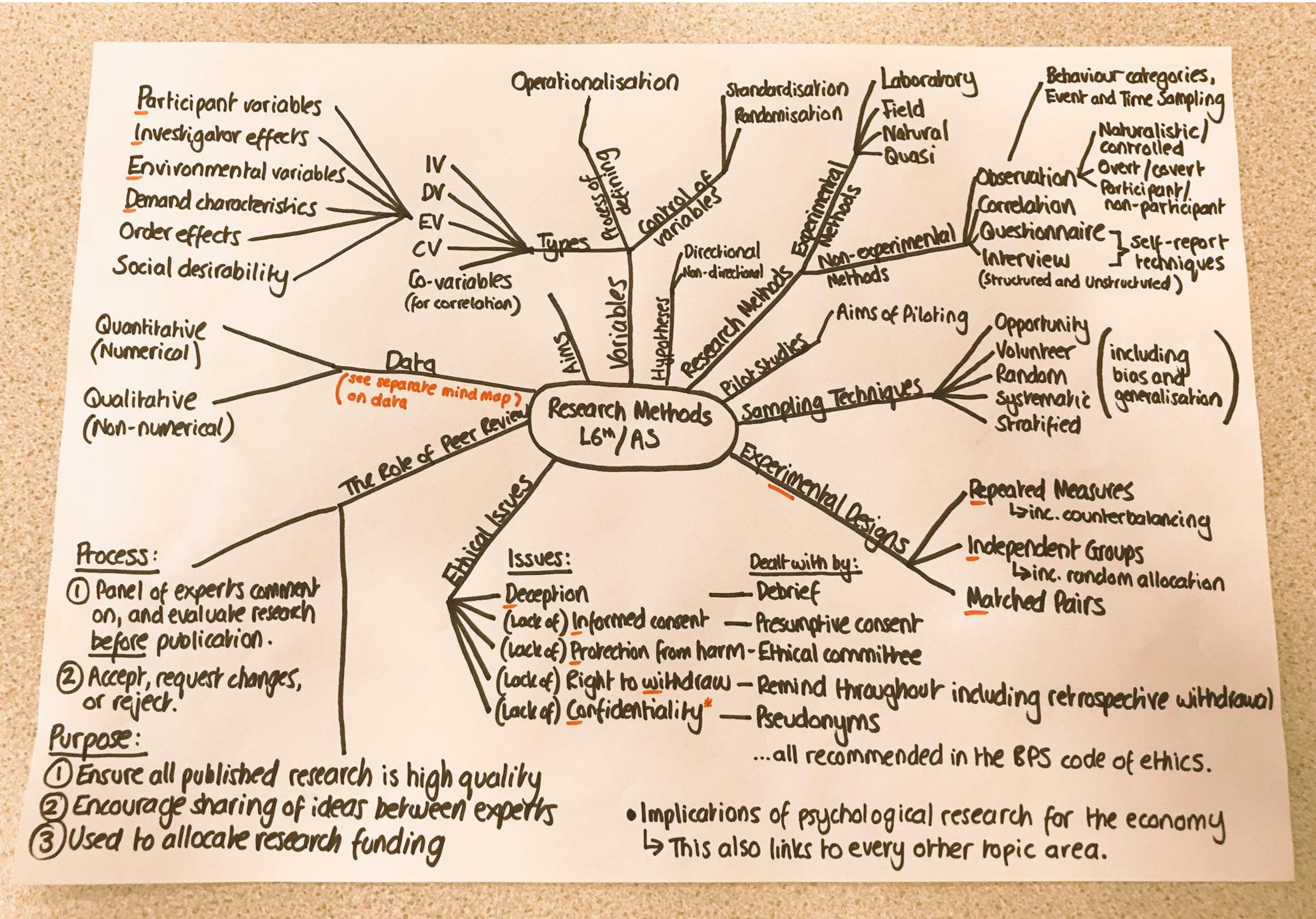
- Standing back from the information given and generate mindmaps
- Checking whether it is completely accurate (evaluate evidences and validation process)
- Checking whether each statement logically flows from what went before (logical reasoning)
- Looking for possible flaws in reasoning, evidence or conclusions
- Being able to see and explain why different people arrive at different conclusions (supported by evidences)
- Being able to argue why one set of opinions, results or conclusions is preferable to another
- Being on guard for literary or statistical devices that encourage the reader to take questionable statements at face value
- Checking for underlying assumptions

Critical analysis when writing

- Clear about the conclusions
- Reasoning out the rationale behind the conclusions
- Use facts and evidences to support reasoning
- Viewing the subject from multiple perspectives/view points
- Writing in analytical style with logical reasoning
- Effective use of mind/concept maps (look for new information and how the existing information structures)

  Mind mapping (or concept mapping) involves writing down a central idea and thinking up new and related ideas which radiate out from the centre

  Use lines, colours, arrows, branches or some other way of showing connections between the ideas generated on your mind map.
- Use reflective practice in writing a critical literature review



Source: https://twitter.com/ALevelPsy/status/1000389710328590336/photo/1

Reflective practice

“reflecting on experience to improve action and professional practice”

"a reflection in a mirror is an exact replica of what is in front of it. Reflection in professional practice, however, gives back not what it is, but what *might* be, an improvement on the original" - Biggs

The cultivation of the capacity to

reflect in action (while doing something)

and

reflect on action (after you have done it)

has become an important feature in research process

Problem Definition

Formulating the Research Problem (RP):

Researcher must identify the problem he wants to study, i.e., he must decide the general area of interest or aspect of a subject-matter that he would like to inquire into.

Problem can be stated in broad general way and then resolve ambiguities if any.

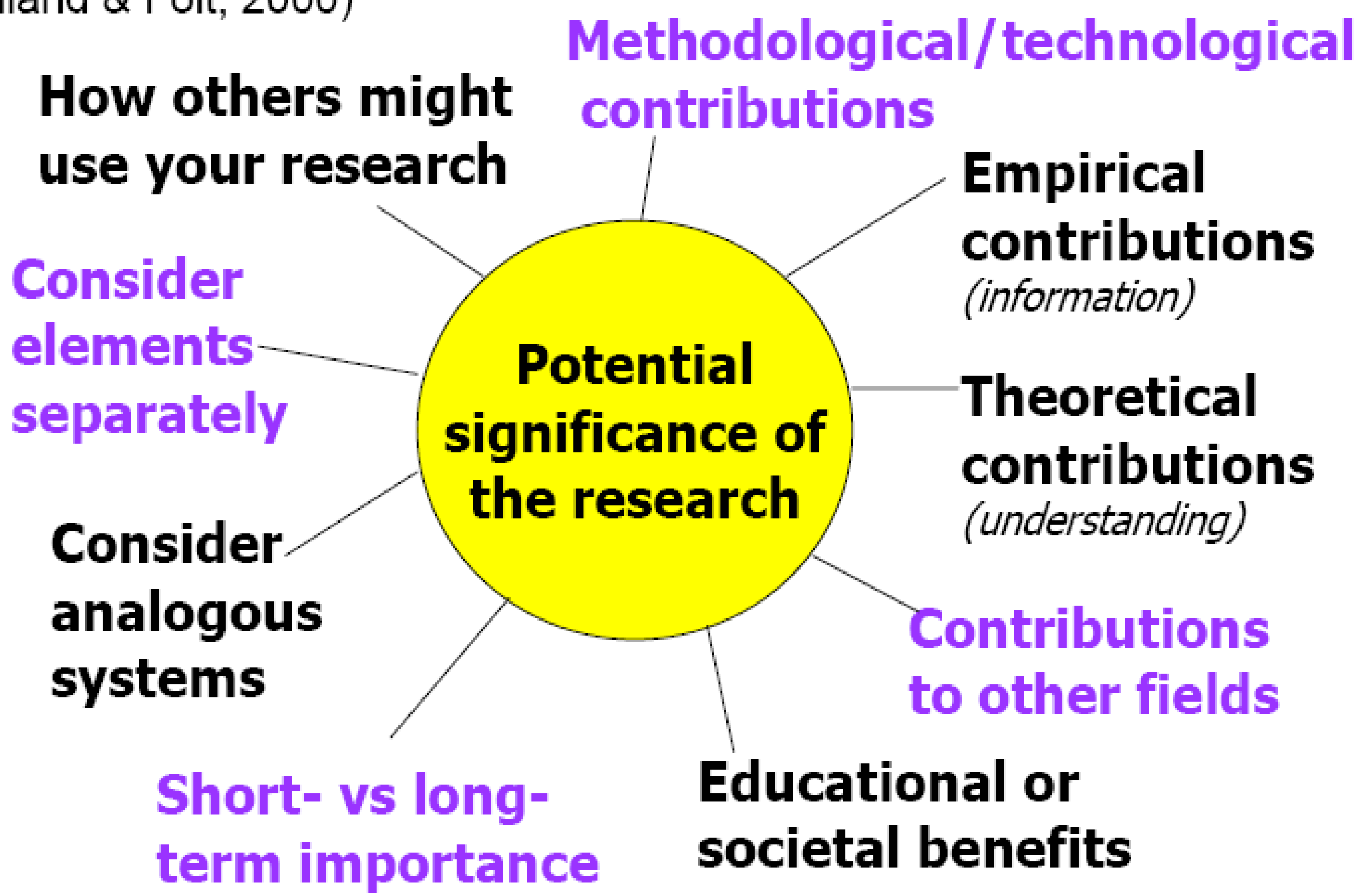
Identification of the problem (in a broader sense) & proper understanding of the problem is the basis of research inquiry (what is the problem, why is it importance and how is it being addressed)

Discussions with colleagues, experts, agencies (ICTA, DEMP, etc.) working in the subject area help to understand the problem

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Significance of research (research impact)

(Friedland & Folt, 2000)



Problem Definition

Formulating the Research Problem (RP):

Examine all available literature to get yourself familiar with the selected problem

Two types of literature:

1. The conceptual literature concerning the concepts and theories
2. Empirical literature consisting of studies made earlier which are similar to the one proposed

Outcome: the knowledge as to what data and other material are available for operational purposes

Problem Definition

It is fruitful to define a problem as any situation where a gap exists between the actual and the desired ideal state.

It is very important to avoid symptoms of problems are defined as the real problem

Problem Definition or Problem Statement:

Is a clear, precise, and succinct statement of the question or issue that is to be investigated with the goal of finding an answer or solution

Research Objectives

Objectives are the goals that are expected to be attained through the research

In other words,

Objectives explain the steps to be taken in order to answer the research question

They clarifie what researcher wants to achieve through research study. It is very important to word them clearly and specifically.

Normally they are listed under two headings:

Main Objective

Sub Objectives

Research Objectives

Main Objective: A statement of the main associations and relationships that you seek to discover or establish (broad in scope and intangible in nature)

Sub Objectives: The specific aspects of the topic that you want to investigate within the main framework of research study (specific, tangible, concrete and validated)

Aspects:

should be worded clearly and unambiguously

each sub objective contains only one aspect of the study

Normally starts with action oriented words or verbs such as 'to determine', 'to find out', 'to ascertain', 'to measure', 'to explore' etc. (discover, establish, determine, access identify, compare, analyze, evaluate, describe, estimate, develop, calculate)

Research Objectives

The way objectives are worded determines how your research is classified (e.g. descriptive, correlational, or experimental)

Objectives determine the type of research design (research methodology)

Crafting objectives

- Objectives Should be SMART (Specific, Measurable, Achievable, Realistic and Time bound)
- Specific: Describe an observable action or achievement and be precise. Relate to rate, number or percentage
  - To answer phone quickly (not an objective)
  - To answer the phone in no more than 4 rings
- Measurable: A monitoring system in place which allows for tracking and recording outcomes. System may be reliable and may be in existence
  - E.g. Grades, Benchmark testing, Likert scale used in surveys
  - To increase income in 2020 by at least 10% over 2019.
- Achievable: likelihood of being success. Objectives need to be set based on the available resources and time
  - To answer the phone in its first ring
  - To answer the phone in no more than 4 rings

Crafting objectives

- **Realistic/Relevant:** your objectives should be result-oriented and in line with main objective (goal)
  - –To complete the SPSS analysis of the survey data by 8th March 2020.
- **Time bound:** Time frame of the objectives need to be specific unless not defined the time frame of the project. In the methodology, time when the measurements are taken, should be clearly defined (starting point, ending point and fixed durations).
  - To increase the average daily attendance rate to over 95%within the next 3 years.

- Note: A methodology is not just a list of research tasks but an argument as to why these tasks add up to the best attack on the problem" (Przeworski& Salomon)

Research Objectives

A study into the cost and quality of home-based care for HIV/AIDS patients and their communities in Zimbabwe, developed at a workshop, for example, had as its general objective:

To explore to what extent community home-based care (CHBC) projects in Zimbabwe provide adequate, affordable and sustainable care of good quality to people with HIV/AIDS, and to identify ways in which these services can be improved.

It was split up in the following specific objectives:

1. To identify the full range of economic, psychosocial, health/nursing care and other needs of patients and their families affected by AIDS.
2. To determine the extent to which formal and informal support systems address these needs from the viewpoint of service providers as well as patients.
3. To determine the economic costs of CHBC to the patient and family as well as to the formal CHBC programmes themselves.
4. To relate the calculated costs to the quality of care provided to the patient by the family and to the family/patient by the CHBC programme.
5. To determine how improved CHBC and informal support networks can contribute to the needs of persons with AIDS and other chronically and terminally ill patients.

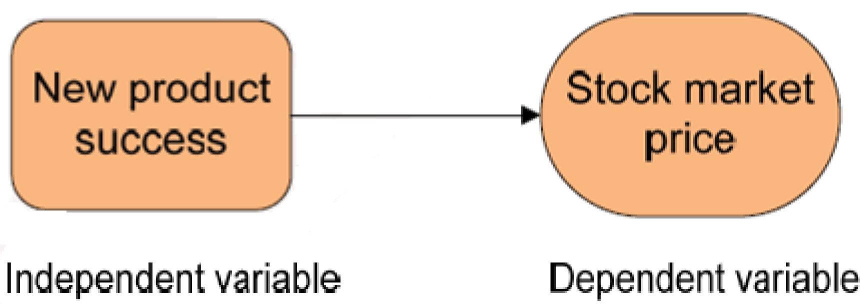
Theoretical Framework

Variables

A variable is anything that can take on differing or varying values. Variables can be discrete or continuous.

Dependent & Independent Variable: An independent variable is one that influences the dependent variable (variable of primary interest to the researcher)

Eg: Research studies indicate that <u>successful new product development</u> has an influence on the <u>stock market price</u> of the company.

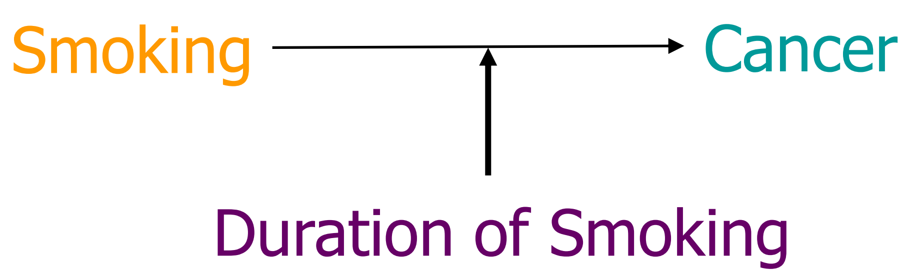


Theoretical Framework

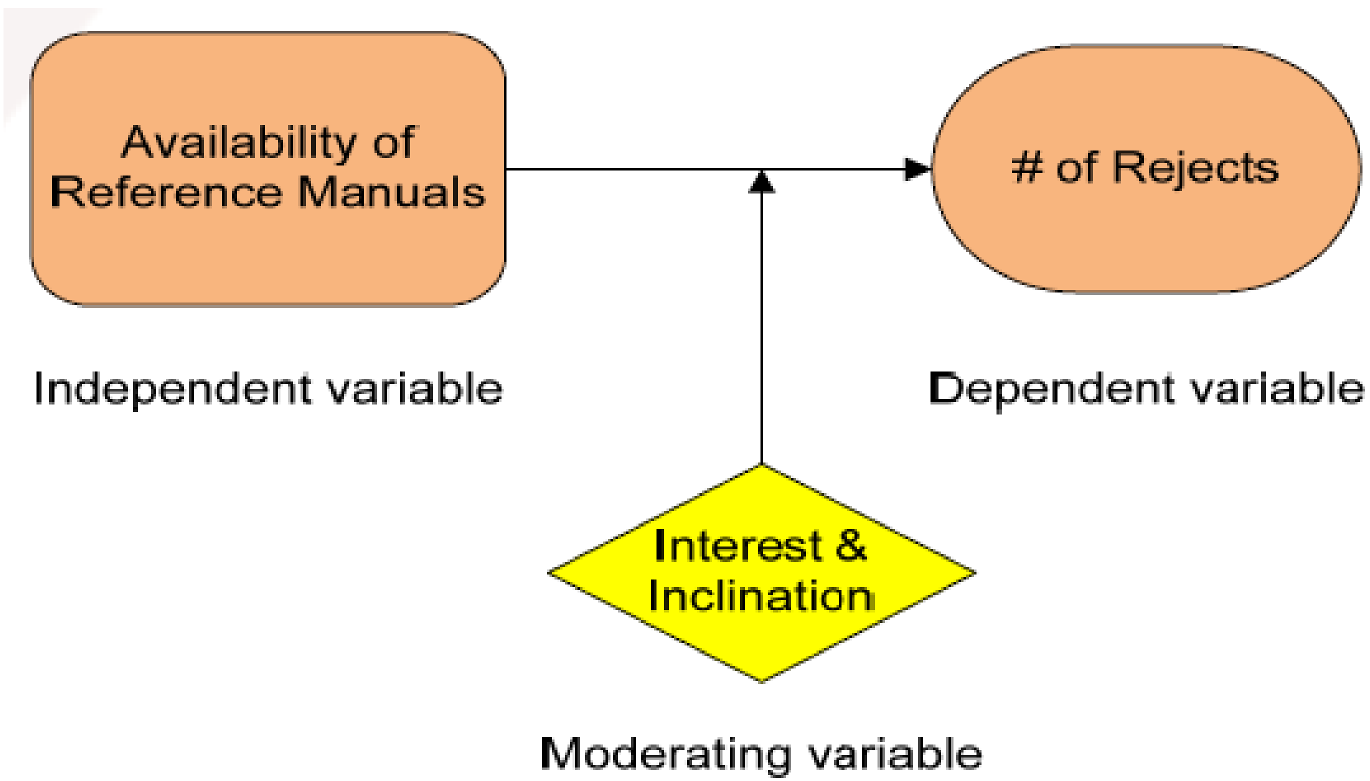
Moderating Variables

A **moderator variable**, commonly denoted as just M, is a third variable that affects the strength of the relationship between a dependent and independent variable. In correlation, a moderator is a third variable that affects the correlation of two variables.

Eg: Smoking is a cause of cancer. The duration of smoking, the age of smoker, dietary habits affects this relationship.



Example - Moderating variable



Theoretical Framework

Intervening Variables

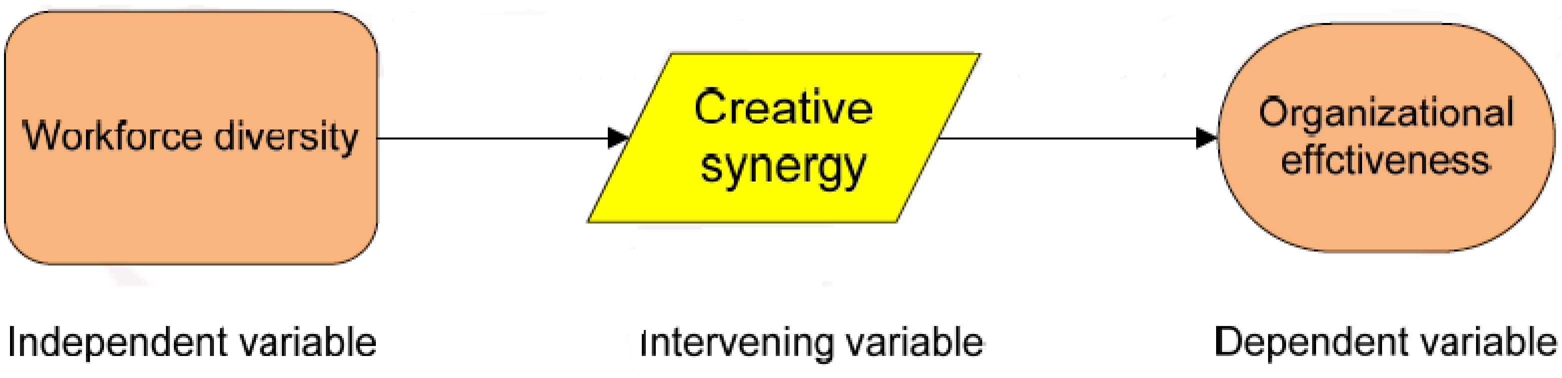
An intervening variable (sometimes called a mediating variable) is a hypothetical variable used to explain causal links between other variables.

Intervening variables cannot be observed in an experiment (that's why they are hypothetical).

For example, there is an association between being poor and having a shorter life span. Just because someone is poor doesn't mean that will lead to an early death, so other hypothetical variables are used to explain the phenomenon.

These intervening variables could include: lack of access to healthcare or poor nutrition.

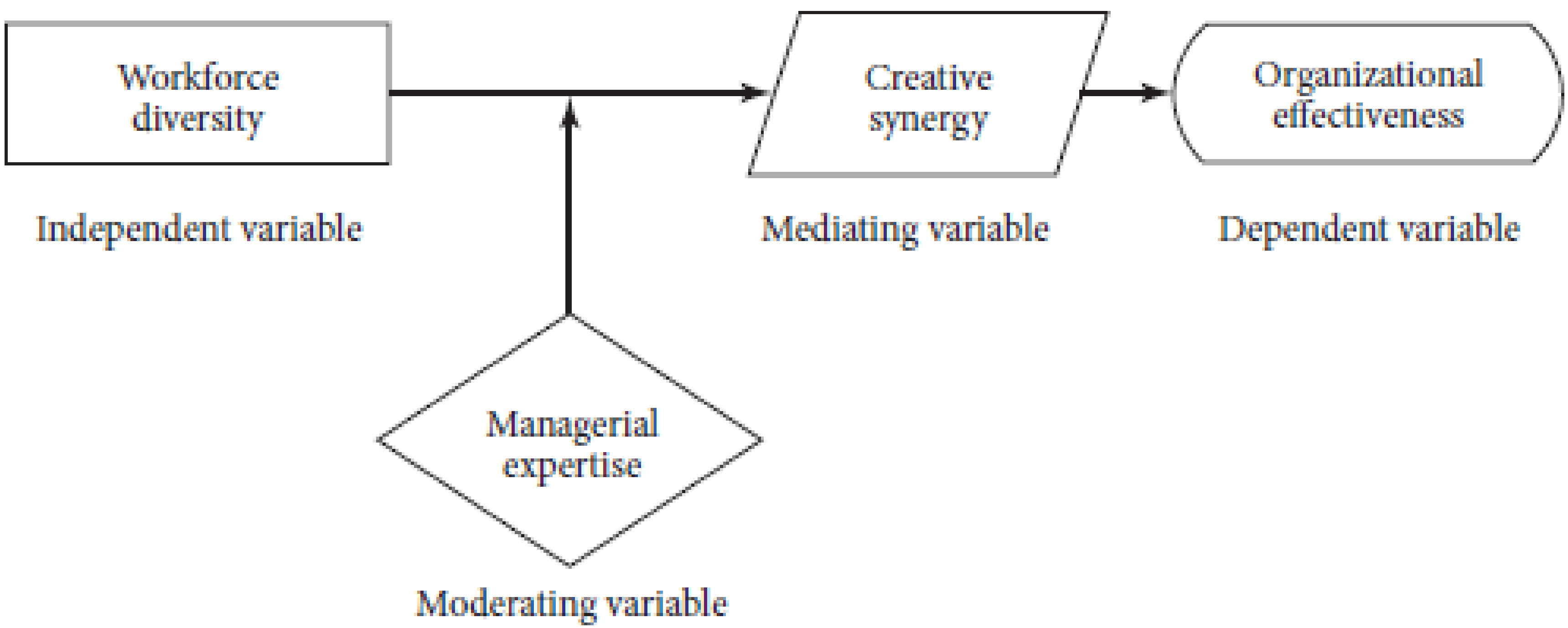
Example – Intervening variable



The *independent variable* workforce diversity influences the *dependent variable* organizational effectiveness, the *intervening variable* that surfaces as a function of diversity in the workforce is creative synergy.

synergy: working together of two things (muscles or drugs for example) to produce an effect greater than the sum of their individual effects

Moderating and mediating variables



Theoretical Framework

A Theoretical Framework (TF) is,

A Conceptual Model of how one theorizes or makes logical sense of the relationships among the several factors that have been identified as important to the problem

A theoretical framework guides your research, determining what things you will measure, and what statistical relationships you will look for.

TF is a logically developed, described, and elaborated network of associations among the variables deemed relevant to the problem situation and identified through such processes as interviews, observations, and literature survey

Theoretical Framework

Five basic features of TF:

1.Variables relevant to the study should be clearly identified and labeled

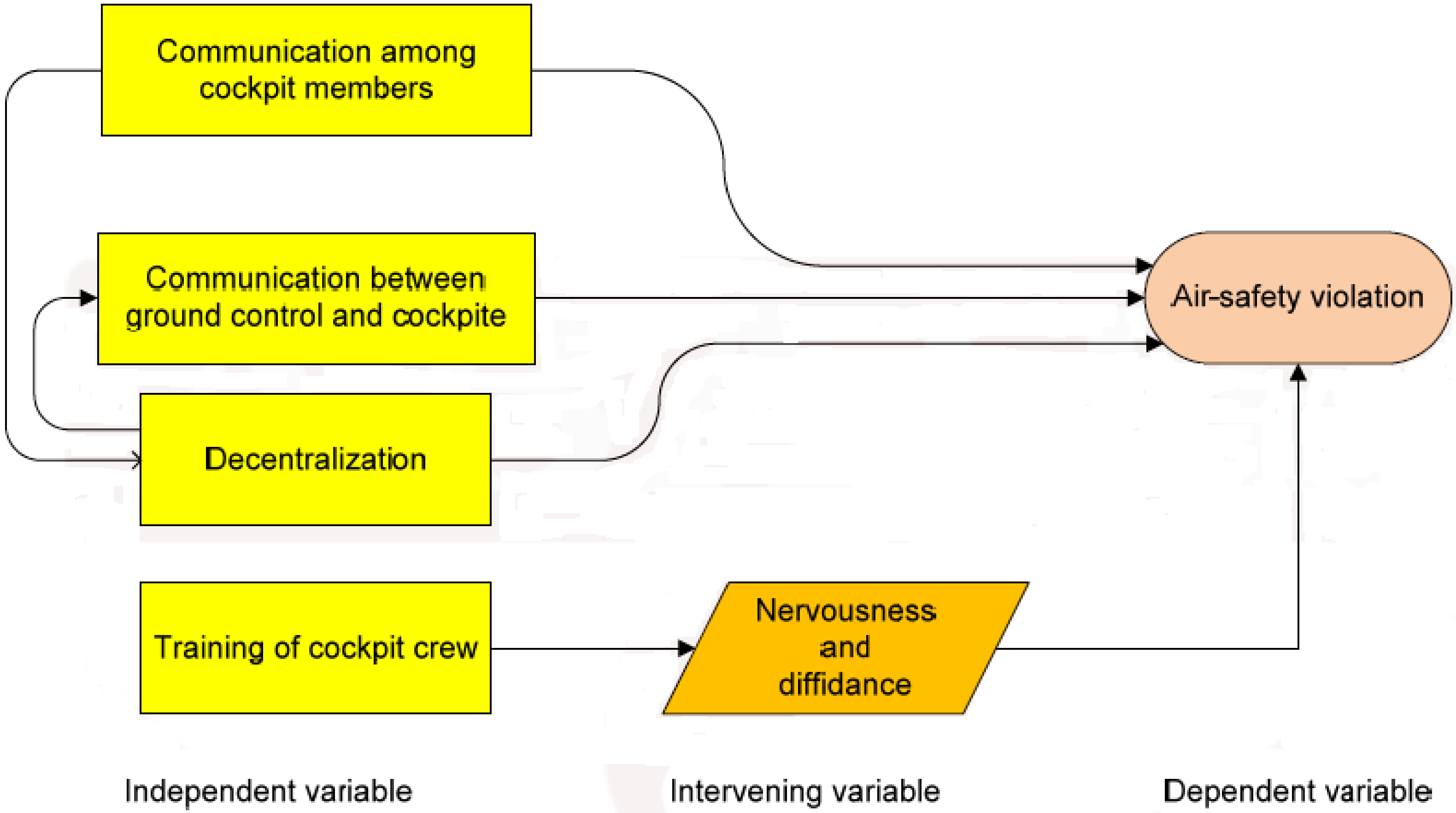
2. Description of how two or more variables are related to one another

3. If nature and direction of the relationship is theorized based on previous research, outcome should be stated

4. Clear explanation of why we would expect these relationships to exist.

5. A schematic diagram of the TF

Schematic diagram of TF



Generation of hypothesis

- **Definition:** A logically conjectured relationship between two or more variables expressed in the form of testable statement
- Statement of hypothesis – Formats
  - Propositions and If-Then Statements:
    - Proposition: Ex. Employees who are more healthy will take sick leave less frequently
    - If-Then: Ex. If employees are more healthy, Then they will take sick leave less frequently
  - Directional and Non-directional:
    - Directional: Ex. The greater the stress experienced in the job, the lower the job satisfaction of employees
    - Non-directional: Ex. There is a relationship between age and job satisfaction

Generation of hypothesis and testing

–Null and Alternative Hypotheses:

- Null Hypotheses: is a proposition that stats a definitive, exact relationship between two variables. The null statement is expressed as no significant relationship between two variables or no significant difference between two groups
- Alternate Hypotheses: is the opposite of null, it is a statement expressing a relationship between two variables or indicating differences between groups

- **Hypothesis testing:** It explains the nature of certain relationships, or establishes the differences among groups or the independence of two or more factors in a situation

- For hypothesis testing various statistical techniques/tests are used

Thank You..!

Research Methods

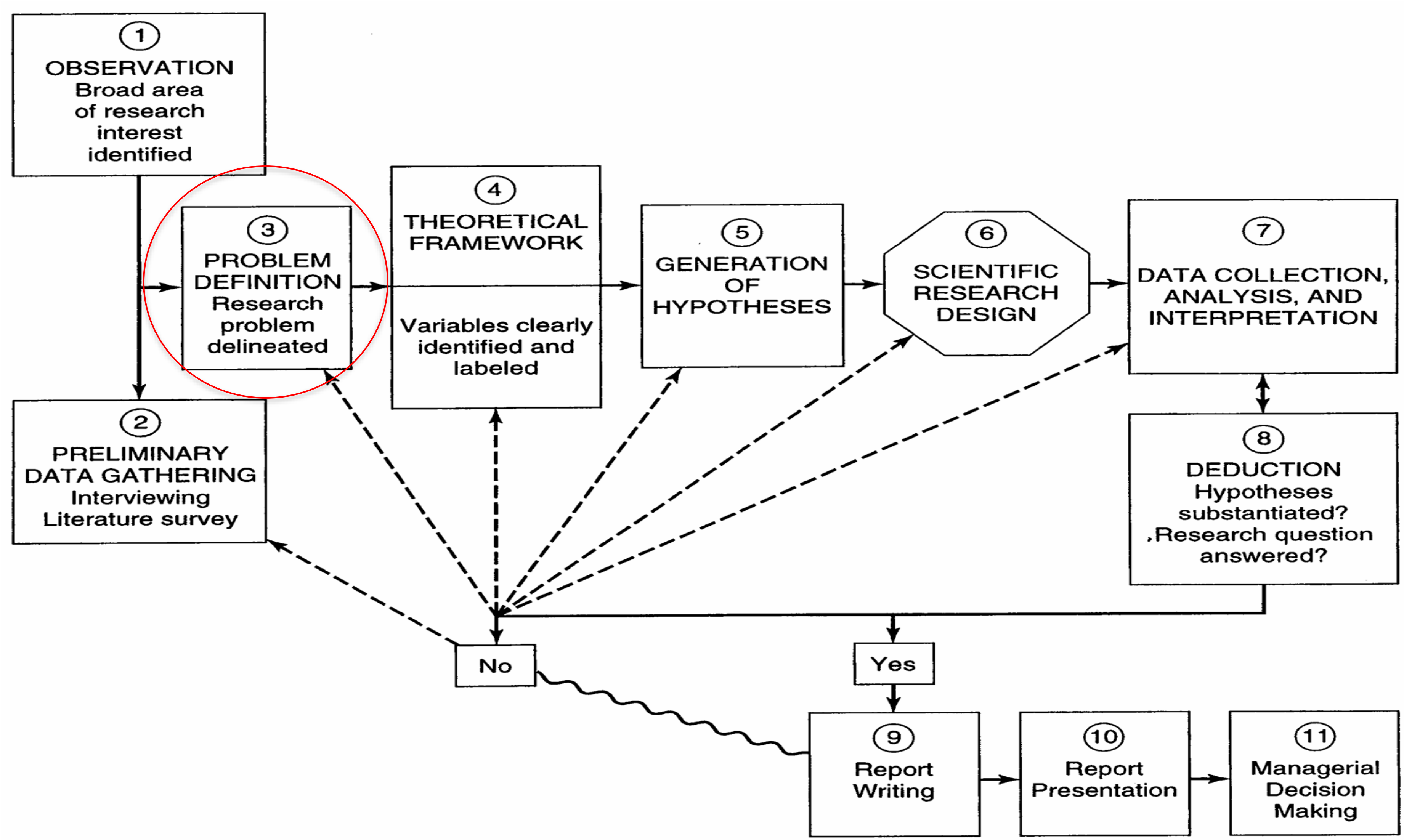
Lecture 3-1: Developing a Research Question

Agenda

- Topic selection
- Refine and define the research question
- Class activity: Writing research question
- Evaluating research question
- Sample research question
- Good question checklist

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Where we are in the steps of research process?



Topic Selection

- **Select a Topic Area of your Choice**
  - Reading Literature tells you what still needs to be done
- **Refine it to formulate the Research Question -RQ**
  - RQ is a problem or area which can be researched through a question or group of sub questions
- **RQ can be answered by**
  - Collecting data relevant to the question
  - Analyzing data by the use of related theory
  - Drawing conclusions so as to answer the questions

Finding your Research Question?

- Choose an appropriate topic or issue
- One that actually can be researched
- Then list all of the questions that you'd like answered yourself.
- Choose the best question, one that is neither too broad nor too narrow.

Narrowing down - Exercise

- **Below is a listing of related topics. Please choose the topic that you consider to be the broadest term.**
  - Transportation
  - Cars
  - Vehicles
  - Ford Explorer

- **Choose the topic that you consider to be the narrowest term.**
  - Plants
  - Flowers
  - Roses
  - Agriculture

Refine the topic

- **Once you decide on your topic, the topic should be written as a thesis statement, or short sentence that:**
  - Defines the main idea of your research
  - States what you will describe or prove in your paper

Developing your research question

I know what general area, but I’m not sure of my research question?

- **A good research question:**
  - Defines the investigation
  - Sets boundaries
  - Provides direction

Research question

- **The investigator must make sure that:**
  - He has a research question
  - The question is clear and specific
  - It reflects the objectives of the study
  - It has no answer by common sense
  - It has no answer in the LITERATURE
  - Finding an answer to the question will solve or at least help in solving the problem to be studied.

How to develop a question

- **Identify key elements.**
- **Ask:**
  - –who, why, where, when, what and how? *(quantitative)*
  - –to what extent …? how do …? what are …? (*qualitative*).

- **Need to get the length of the question right, i.e. a balance between being brief and getting in all the necessary details.**

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Your turn – Can the topic be researched

- **Question A:**
  Does McDonald's or Burger King make a better burger?

- **Question B:**
  Is there a link between hours of television viewing and violent behavior in children aged 8-14?

Check with you - Researchability

- **Question A:**
  Does McDonald's or Burger King make a better burger?
  – not researchable as it is worded, no concrete meaning. What does "better" mean? Better in terms of nutrition? Better tasting? Better value? Fewer calories? Better for making your kids happy?

- **Question B:**
  Is there a link between hours of television viewing and violent behavior in children aged 8-14?
  – researchable. You'd have to sift through a lot of information, both pro and con, valid and invalid, in order to choose the best information to answer the research question and support your own point of view, but the point is that there is at least enough information to sift through.

Your turn: Scope is too narrow or too broad

- **A: What marketing strategies does the Coca-Cola company currently apply?**
- **B:What is the Coca-Cola company's future marketing plan?**
- **C: What marketing strategies has the Coca-Cola company used in the past?**

Check with you - scope

- **A:What marketing strategies does the Coca-Cola company currently apply?**
  - Good RQ. Answer to this question may include observation of print, television, and radio advertisements as well as research into various current marketing theories and strategies.

- **B:What is the Coca-Cola company's future marketing plan?**
  - B is very broad as well as being unresearchable--it's unlikely that Coca-Cola personnel will reveal their marketing plan.

- **C:What marketing strategies has the Coca-Cola company used in the past?**
  - may be too broad as well, since "the past" covers a lot of time, especially since the Coca-Cola company was incorporated in 1919.

Writing research questions

It is important that these questions establish the link between your study and other research that has preceded it; they should also clearly show the relationship of your research to your field of study.

- **Research questions should guide:**
  - The formulation of your research plan
  - The aims and objectives of your study
  - The literature review
  - Decisions about the kind of research design to employ
  - Decisions about what data to collect and from whom
  - The analysis of data
  - The writing up of the work
  - The research direction - *i.e.* addressing the questions should stop you from going off in unnecessary directions and tangents.

Class activity

- Write a research question for your selected topic.

Evaluate your own research question

1. **Does the question deal with a topic/issue of interests ?**

2. **Is the question easily and fully researchable?**

3. **What type of information do I need to answer the research question?**

   Eg. "What impact has deregulation had on commercial airline safety?," requires

   – statistics on airline crashes before and after

   – statistics on other safety problems before and after - information about maintenance practices before and after

   – information about government safety requirements before and after

Evaluate your own research question

4. **Is the scope of this information reasonable?**

   (can I research all marketing campaigns conducted over a span of 10 years?)
5. **Given the type and scope of the information that I need, is my question too broad, too narrow, or okay?**
6. **What sources will have the type of information that I need to answer the research question**

   (journals, books, Internet resources, government documents, people)?
7. **Can I access these sources?**
8. **Given my answers to the above questions, do I have a good quality research question that I actually will be able to answer by doing research?**

Sample research question 1

Title : Improving Nanjing University English Majors’ Reading Performance with Self-Access Language Learning (SALL)

Aims: Aims of the research were to study the impact of self-access learning on students’ performance and to explore their attitudes towards SALL.

Questions:

1. Does second-year students’ reading performance from two interventions (directed and self-directed) differ from each other?
2. To what extent do the students’ attitudes towards the self-access program enhance their performance? Do these attitudes promote the integration of SALL pedagogy in Nanjing University?

Case study: develop research questions

CAA Airlines carries out charter and regular flights to medium-haul destinations – such as the Mediterranean, North Africa and the Red Sea – and to long-haul destinations such as the Caribbean. Today, CAA ' s fleet consists of three (new) Boeing 737-800s and four (outdated) Boeing 767-300s. Because the Boeing 767s are rather outdated they need more maintenance than the average airplane. Despite an intensive maintenance program, these planes have a lot of technical problems. Consequently, the long-haul fleet of CAA has needed to deal with a lot of delays recently. New long-haul planes have been ordered, but these planes will not be delivered before 2019. This means that more delays will inevitably occur. This may translate into much frustration among airline passengers, to switching behavior, and to negative word-of mouth communication.

These feelings and behaviors of consumers may eventually have negative effects on the performance and the profitability of the firm.

Prior research has claimed that service waits can be controlled by two techniques: operations management and management of perceptions. For CAA Airlines it is very difficult to obtain "zero defects" (no delays). Hence, this project will focus on managing the perceptions of the wait experience: because CAA Airlines cannot control the actual amount of delays and the duration, the company must focus on managing the customers ' perception of the waiting experience.

The purpose of this study is twofold:

(1) to identify the factors that influence the passengers ' waiting experience and

(2) to investigate the possible impact of waiting on customer satisfaction and service evaluations.

Case study: Research questions

Therefore, this project focuses on the following research questions:

1. What are the factors that affect the perceived waiting experience of airline passengers and to what extent do these factors affect the perception of waiting times?
2. What are the affective consequences of waiting and how does affect mediate the relationship between waiting and service evaluations?
3. How do situational variables (such as filled time) influence customer reactions to the waiting experience?

Clarity

- **The question(s) should be answerable**
  i.e can be illuminated or addressed by your methodology

- **The question should be intelligible to the reader who may not be an 'expert' in your topic (understandable)**

- **The questions should offer the prospect of making an 'original contribution to knowledge' in some way**

- **Are the terms clearly defined?**

- **Are the questions precise?**

source: Ingrid Lunt.

Empirical focus

- **Require that you generate data to answer question**

- **Lead you to determine methods of enquiry and data collection**
  - different types of questions will lead to different approaches to research and methods of data collection,

Significance

- **Is there a clear rationale for the question?**
- **So what?**
- **Does this question matter?**
- **Why is it of interest and to whom?**

Good question checklist

- **Is the question right for me?**
  - Will the question hold my interest?
  - Can I manage any potential biases/subjectivities I may have?
- **Is the question right for the field?**
  - Will the findings be considered significant?
  - Will it make a contribution?
- **Is the question well articulated?**
  - Are the terms well-defined?
  - Are there any unchecked assumptions?

Good question checklist

- **Is the question doable?**
  - Can information be collected in an attempt to answer the question?
  - Do I have the skills and expertise necessary to access this information? If not, can the skills be developed?
  - Will I be able to get it all done within my time constraints?
  - Are costs likely to exceed my budget?
  - Are there any potential ethics problems?
- **Does the question get the tick of approval from those in the know?**
  - Does my supervisor think I am on the right track?
  - Do 'experts' in the field think my question is relevant/ important/ doable?

Thank You..!

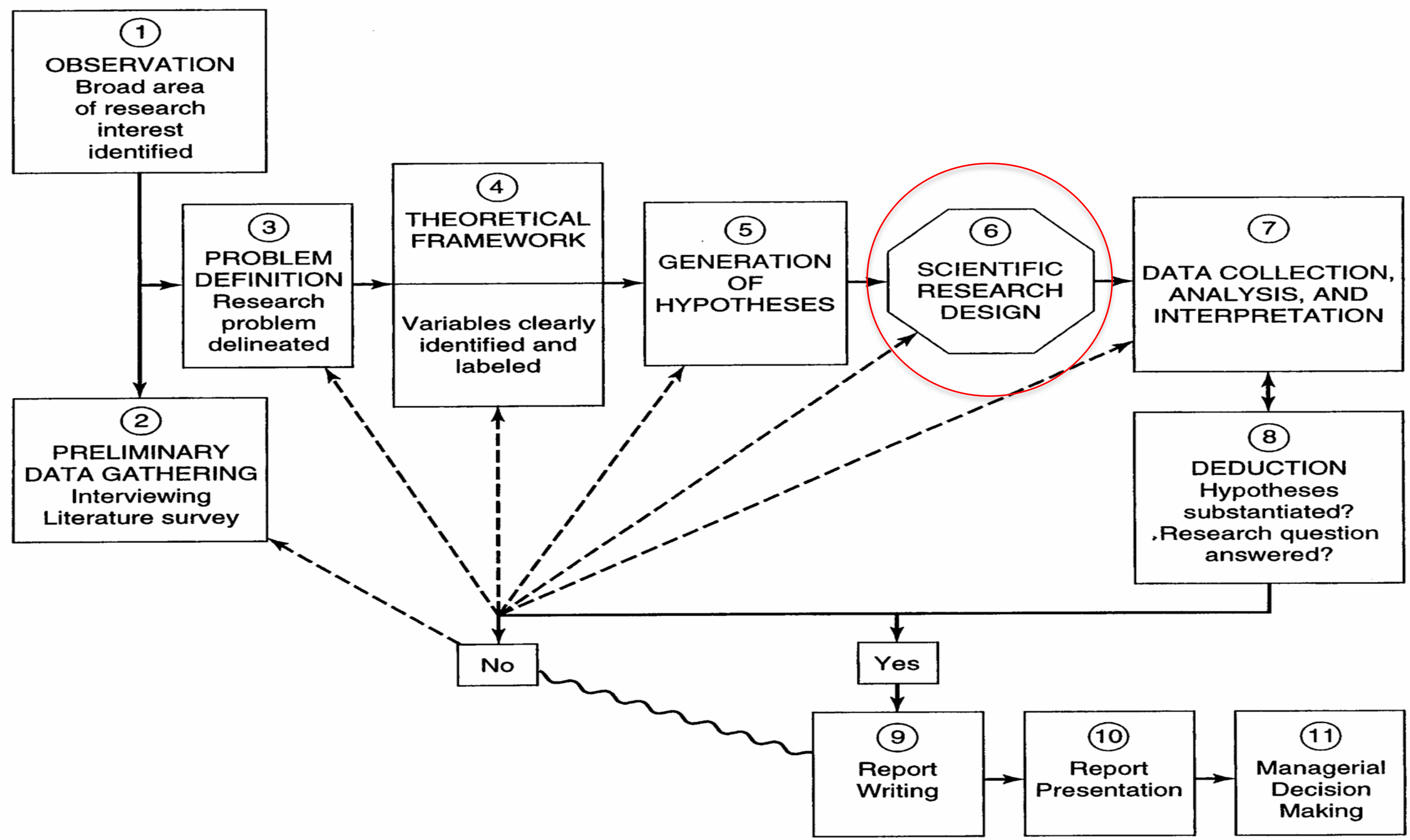
Research Methods

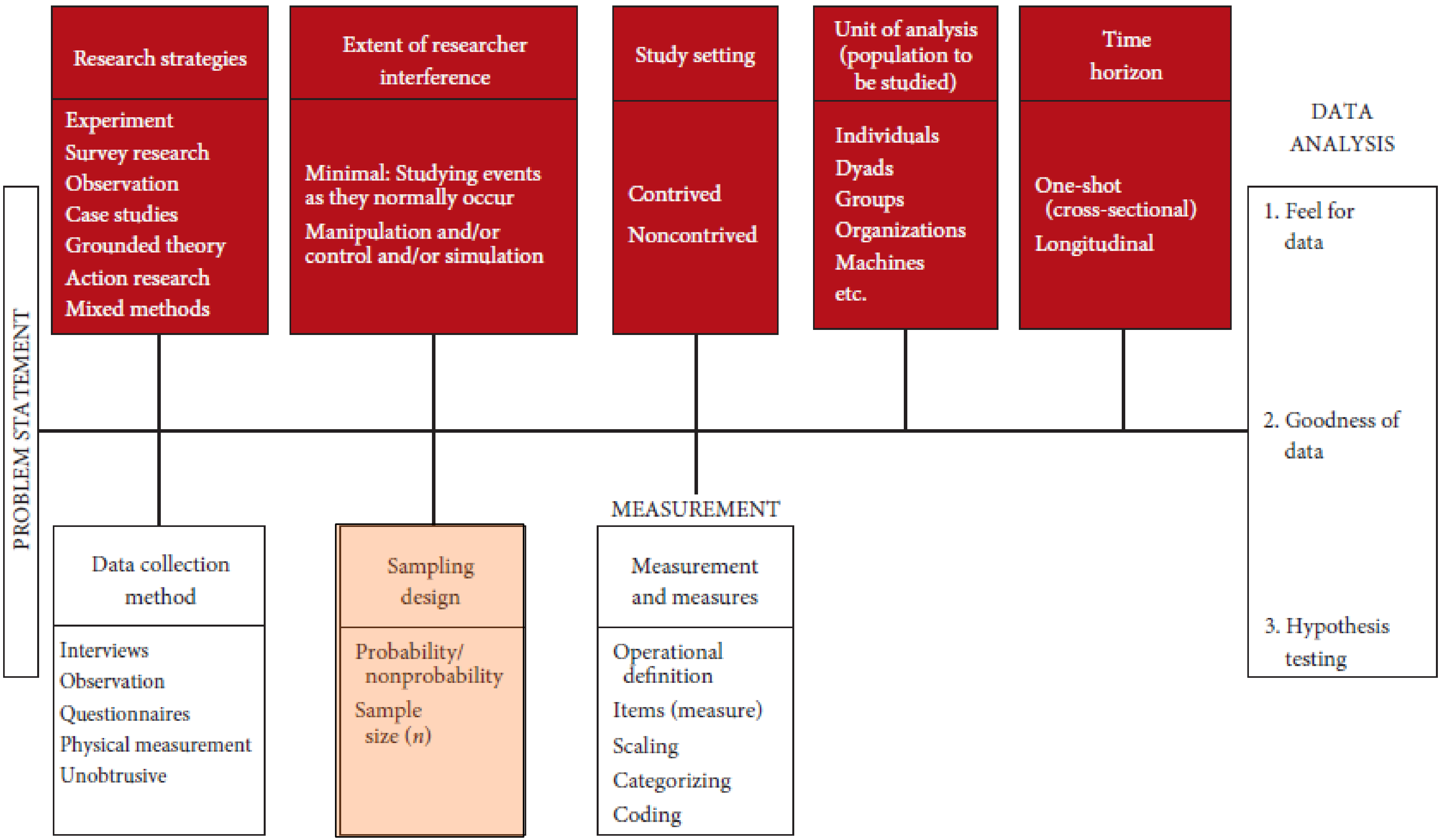
Lecture 4-1: *Sampling Design*

Agenda

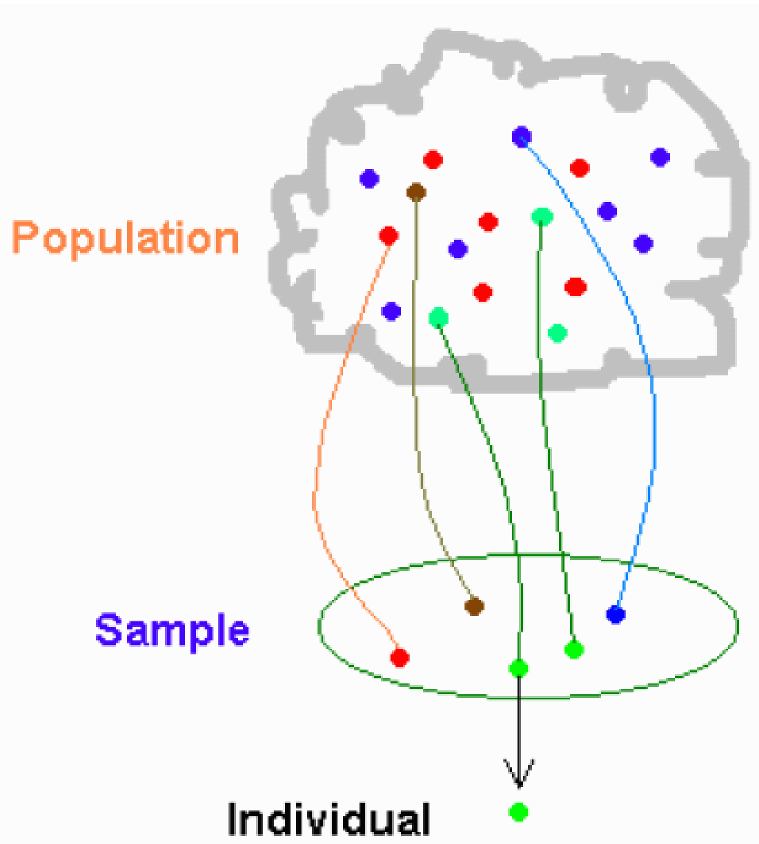
- Population and sample
- Important definitions
- What is sampling
- Properties of sampling and sampling error
- Sample vs. census
- Sampling design process
- Classifications of sampling techniques
- Probabilistic and non-probabilistic sampling
- Confidence, precision and sample size

Where we are in the steps of research process?




Research strategies
Experiment
Survey research
Observation
Case studies
Grounded theory
Action research
Mixed methods
Extent of researcher interference
Minimal: Studying events as they normally occur
Manipulation and/or control and/or simulation
Study setting
Contrived
Noncontrived
Unit of analysis (population to be studied)
Individuals
Dyads
Groups
Organizations
Machines
etc.
Time horizon
One-shot (cross-sectional)
Longitudinal
DATA ANALYSIS
1. Feel for data
2. Goodness of data
3. Hypothesis testing
PROBLEM STATEMENT
Data collection method
Interviews
Observation
Questionnaires
Physical measurement
Unobtrusive
Sampling design
Probability/ nonprobability
Sample size (n)
MEASUREMENT
Measurement and measures
Operational definition
Items (measure)
Scaling
Categorizing
Coding

Population and sample



- Population: All the individuals of interest, about which we want to make some statement or Refers to the entire group of people, events, or things of interest that the researcher wishes to investigate
  - Example: all farmers in a village / district / province / country
- Some populations are more hypothetical
  - All rice crops that could be produced in a defined area
- Sample: The set of individuals actually observed
- Sample is a <u>subset of the population</u>
  - Example: twenty selected farmers, twenty rice crops actually produced

Important terms

- **Element :** An element is a single member of the population
- **Census:** a study of all individual elements in the population
- **Population Frame :** PF is a listing of all the elements in the population from which the sample is drawn
- **Sampling:** is a process of selecting a small number of items of the population to make conclusions regarding the whole population
- **Sample:** is a subset of or sub group of a large population
- **Subject :** A subject is a single member of the sample

What is sampling

- Sampling is the process of selecting a sufficient number of elements from the population

- Sampling is done to understand and then generalize properties or characteristics of the population

Properties of sampling

- **Representativeness of Samples**
  - The most important aspect of sampling is, sample should replicate the properties or characteristics of population from which it is drawn
  - Sample has to be chosen in a scientific way, to assure that the sample statistic is fairly close to the population parameters

- **Normality of Distributions**
  - Attributes and characteristics of the population are generally normally distributed
  - In sampling we have to ensure that the sample subjects not chosen from the extremes, but are truly representative of the properties of the population

Sampling error

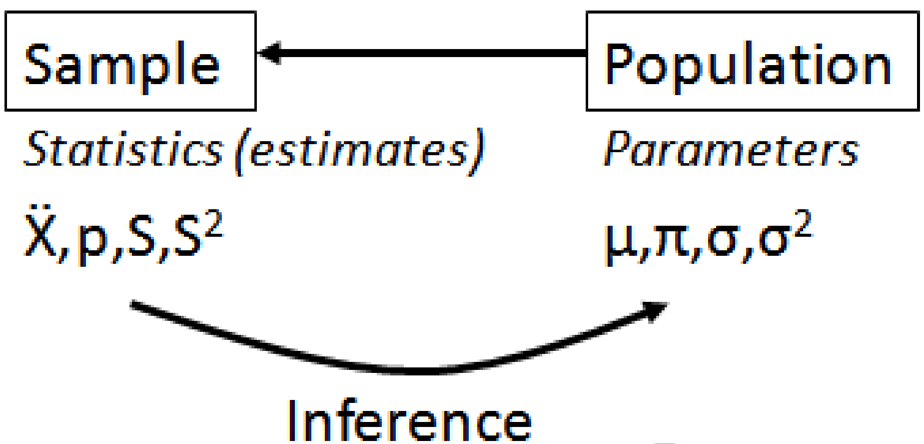
- **Estimates from samples are almost never equal to true values, and estimates from different samples differ among themselves**

- **To quantify this we define the concept of sampling error**
  - The amount by which an estimate of some population parameter computed from a sample deviates from the true value of that parameter for the population.
  - Eg:-surveys carried out to predict the election results

- **Of course we usually don't know this (since we don't know the true value)**

Reasons for sampling

- Sampling is done when it is practically impossible to collect data from, or test, or examine every element
- Sometimes it is possible but not worth in terms of time, cost, and other human resources
- In a few cases, it would be impossible to use the entire population to gain knowledge about, or test something
- Process can be destructive



Sample vs. census

Type of Study	Conditions Favoring the Use of Sample	Census
1. Budget	Small	Large
2. Time available	Short	Long
3. Population size	Large	Small
4. Variance in the characteristic	Small	Large
5. Cost of sampling errors	Low	High
6. Cost of nonsampling errors	High	Low
7. Nature of measurement	Destructive	Nondestructive
8. Attention to individual cases	Yes	No

Sampling design process

Define the Population

↓

Determine the Sampling Frame

↓

Select Sampling Technique(s)

↓

Determine the Sample Size

↓

Execute the Sampling Process

Define the target population

- The target population is the collection of elements or objects that possess the information sought by the researcher and about which inferences are to be made.

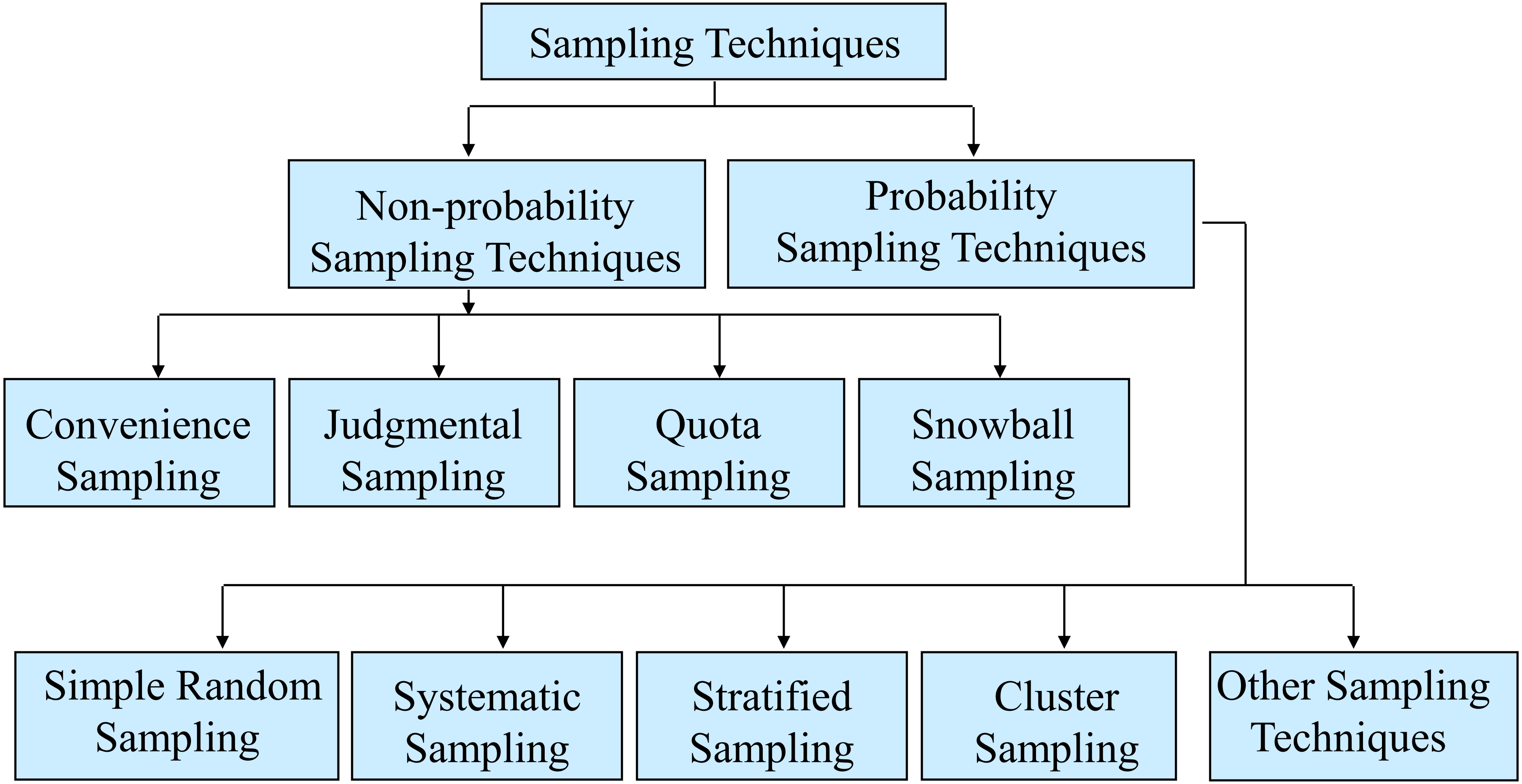
- The target population should be defined in terms of elements, sampling units, extent, and time
  - An **element** is the object about which or from which the information is desired, e.g., the respondent
  - A **sampling unit** is an element, or a unit containing the element, that is available for selection at some stage of the sampling process
  - **Extent** refers to the geographical boundaries
  - **Time** is the time period under consideration

Qualitative factors in determining the sample size

- the importance of the decision
- the nature of the research
- the number of variables
- the nature of the analysis
- sample sizes used in similar studies
- incidence rates
- completion rates
- resource constraints

Type of Study	Minimum Size	Typical Range
Problem identification research (e.g. market potential)	500	1,000-2,500
Problem-solving research (e.g. pricing)	200	300-500
Product tests	200	300-500
Test marketing studies	200	300-500
TV, radio, or print advertising (per commercial or ad tested)	150	200-300
Test-market audits	10 stores	10-20 stores
Focus groups	2 groups	4-12 groups

Classification of sampling techniques



Simple Random Sampling

- **Each element in the population has a known and equal probability of selection.**

- **Each possible sample of a given size (n) has a known and equal probability of being the sample actually selected**

- **This implies that every element is selected independently of every other element.**

  - **Advantages:** High generalizability of findings
  - **Disadvantage:** larger errors for some sample size than with stratified sampling; requires sampling frame at start.

Simple Random Sampling

1. Select a suitable sampling frame
2. Each element is assigned a number from 1 to N (pop. size)
3. Generate n (sample size) different random numbers between 1 and N
4. The numbers generated denote the elements that should be included in the sample

Systematic Sampling

- **Systematic sampling design involves drawing every *n*th element in the population starting with a randomly chosen element between *1* and *n***
- **If the ordering of the elements produces a cyclical pattern, systematic sampling may decrease the representativeness of the sample**
- **For example,** there are 100,000 elements in the population and a sample of 1,000 is desired. In this case the sampling interval, i, is 100. A random number between 1 and 100 is selected. If, for example, this number is 23, the sample consists of elements 23, 123, 223, 323, 423, 523, and so on

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Systematic Sampling

1. Select a suitable sampling frame
2. Each element is assigned a number from 1 to N (pop. size)
3. Determine the sampling interval i:i=N/n. If i is a fraction, round to the nearest integer
4. Select a random number, r, between 1 and i, as explained in simple random sampling
5. The elements with the following numbers will comprise the systematic random sample: r, r+i,r+2i,r+3i,r+4i,...,r+(n-1)i

Stratified sampling

- Two step process:
  - stratification and segregation,
  - followed by random selection of subjects from each stratum.
- The population is first divide into mutually exclusive groups that are relevant, appropriate, and meaningful in the context of the study
- The strata should be mutually exclusive and collectively exhaustive
- In that every population element should be assigned to one and only one stratum and no population elements should be omitted
- The elements within a stratum should be as homogeneous as possible, but the elements in different strata should be as heterogeneous as possible

Stratified sampling

- A major objective of stratified sampling is to increase precision without increasing cost
- In proportionate stratified sampling, the size of the sample drawn from each stratum is proportionate to the relative size of that stratum in the total population.
- In disproportionate stratified sampling, the size of the sample from each stratum is proportionate to the relative size of that stratum and to the standard deviation of the distribution of the characteristic of interest among all the elements in that stratum.

Stratified Sampling

1. Select a suitable frame
2. Select the stratification variable(s) and the number of strata, H
3. Divide the entire population into H strata. Based on the classification variable, each element of the population is assigned to one of the H strata
4. In each stratum, number the elements from 1 to N_h (the pop. size of stratum h)
5. Determine the sample size of each stratum, n_h, based on proportionate or disproportionate stratified sampling, where

$$\sum_{h=1}^{H} n_h = n$$

6. In each stratum, select a simple random sample of size n_h

Cluster sampling

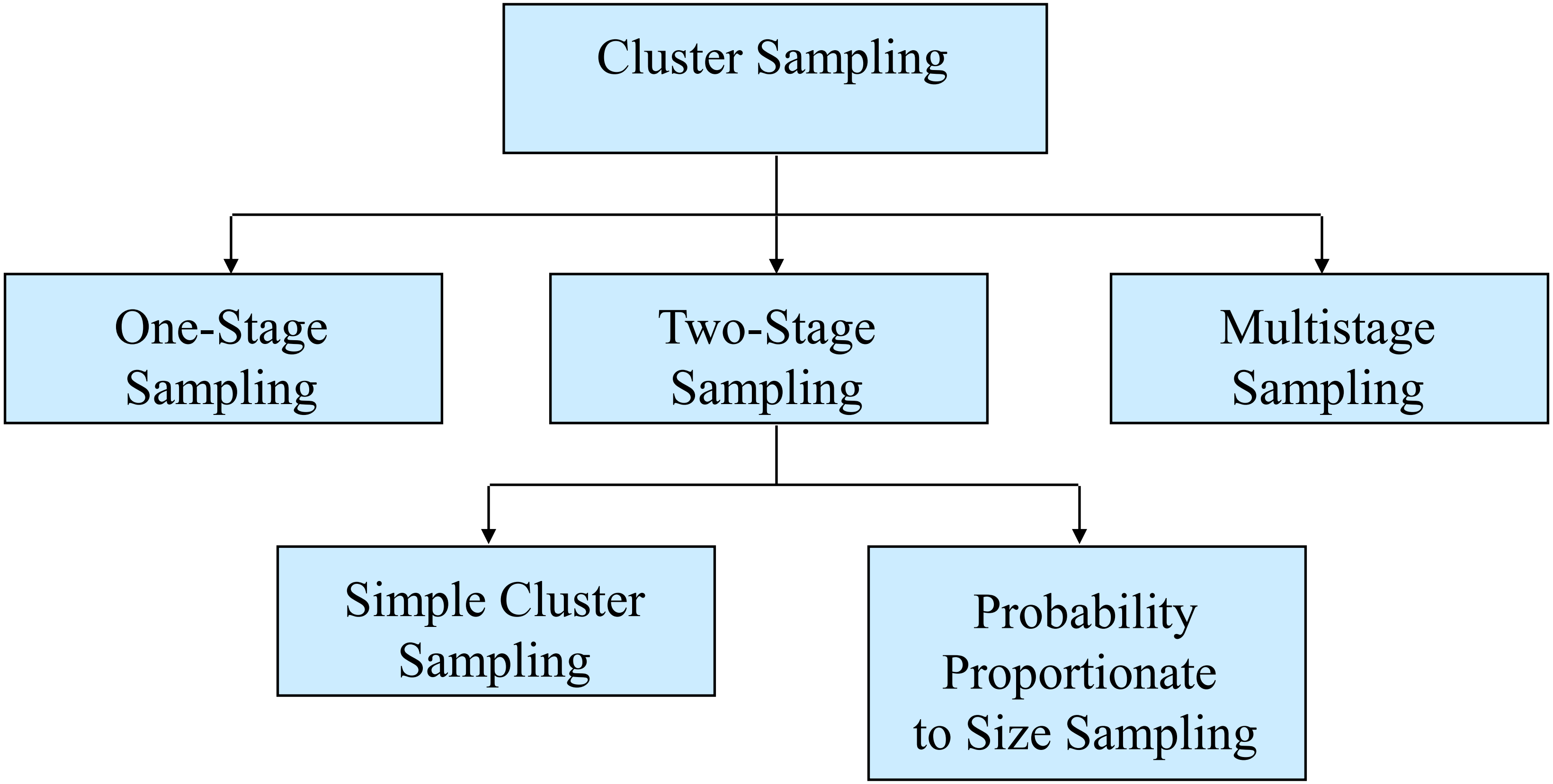
- The target population is first divided into mutually exclusive and collectively exhaustive subpopulations, or clusters
- Then a random sample of clusters is selected, based on a probability sampling technique such as simple random sampling
- For each selected cluster, either all the elements are included in the sample (one-stage) or a sample of elements is drawn probabilistically (two-stage)

Cluster sampling

- Elements within a cluster should be as heterogeneous as possible, but clusters themselves should be as homogeneous as possible. Ideally, each cluster should be a small-scale representation of the population

- In probability proportionate to size sampling, the clusters are sampled with probability proportional to size. In the second stage, the probability of selecting a sampling unit in a selected cluster varies inversely with the size of the cluster

Types of cluster sampling



SLIIT
COMPUTING | BUSINESS | ENGINEERING

Cluster Sampling

1. Assign a number from 1 to N to each element in the population
2. Divide the population into C clusters of which c will be included in the sample
3. Calculate the sampling interval i, i=N/c (round to nearest integer)
4. Select a random number r between 1 and i, as explained in simple random sampling
5. Identify elements with the following numbers: r,r+i,r+2i,... r+(c-1)i
6. Select the clusters that contain the identified elements
7. Select sampling units within each selected cluster based on SRS or systematic sampling
8. Remove clusters exceeding sampling interval i. Calculate new population size N*, number of clusters to be selected C*= C-1, and new sampling interval i*.

Convenient sampling

- **Convenience sampling attempts to obtain a sample of convenient elements. Often, respondents are selected because they happen to be in the right place at the right time**
  - use of students, and members of social organizations
  - mall intercept interviews without qualifying the respondents
  - department stores using charge account lists
  - “people on the street” interviews

Judgmental Sampling

- **Judgmental sampling is a form of convenience sampling in which the population elements are selected based on the judgment of the researcher**
  - test markets
  - purchase engineers selected in industrial marketing research
  - bellwether precincts selected in voting behavior research
  - expert witnesses used in court

SLIIT
COMPUTING | BUSINESS | ENGINEERING

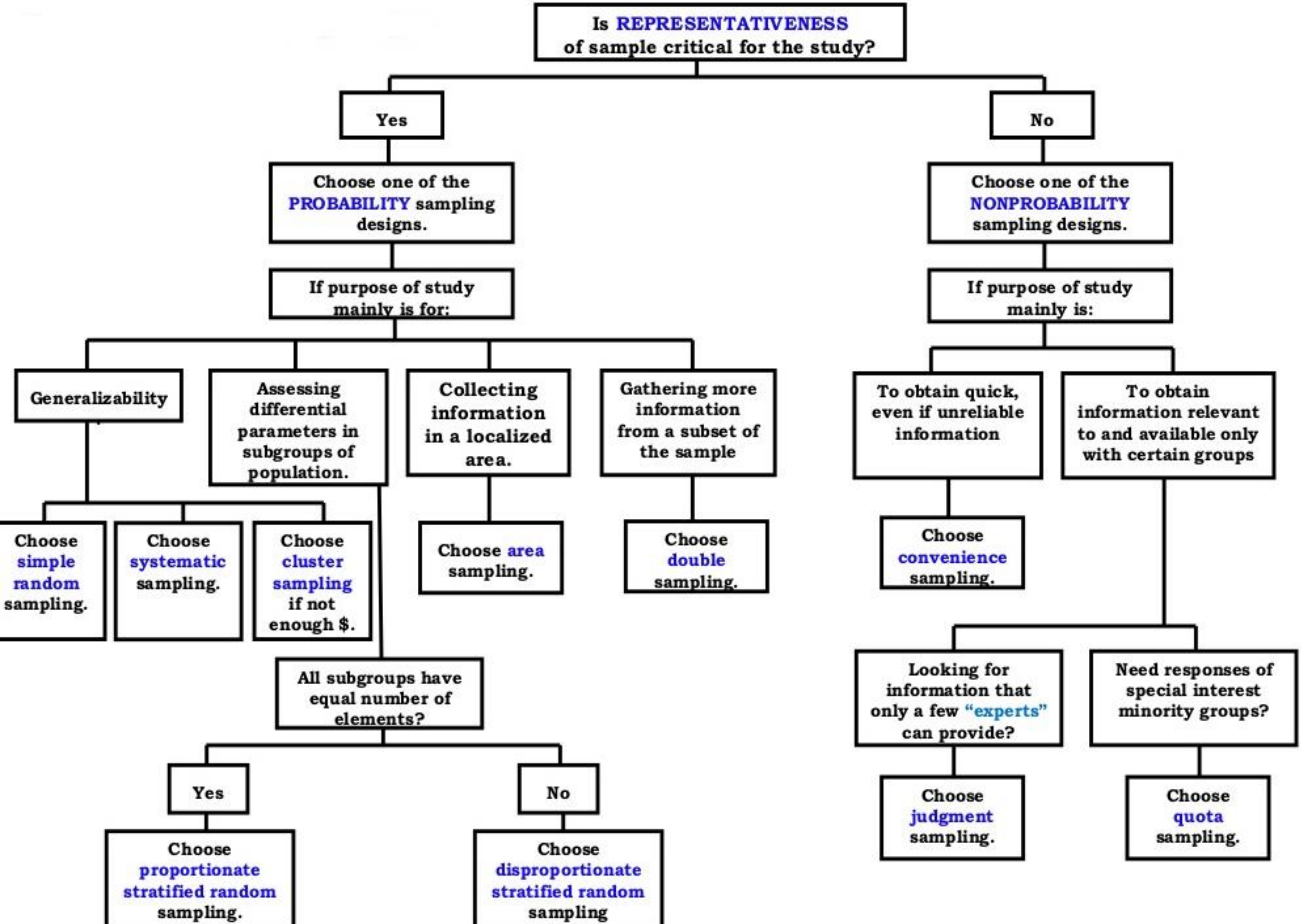
Quota Sampling

- **Quota sampling may be viewed as two-stage restricted judgmental sampling.**
  - The first stage consists of developing control categories, or quotas, of population elements.
  - In the second stage, sample elements are selected based on convenience or judgment.

	Population composition	Sample composition	
Control Characteristic	Percentage	Percentage	Number
Sex			
Male	48	48	480
Female	52	52	520
	100	100	1000

Snowball Sampling

- In snowball sampling, an initial group of respondents is selected, usually at random
  - After being interviewed, these respondents are asked to identify others who belong to the target population of interest
  - Subsequent respondents are selected based on the referrals


Is REPRESENTATIVENESS of sample critical for the study?
Yes
No
Choose one of the PROBABILITY sampling designs.
Choose one of the NONPROBABILITY sampling designs.
If purpose of study mainly is for:
If purpose of study mainly is:
Generalizability
Assessing differential parameters in subgroups of population.
Collecting information in a localized area.
Gathering more information from a subset of the sample
To obtain quick, even if unreliable information
To obtain information relevant to and available only with certain groups
Choose simple random sampling.
Choose systematic sampling.
Choose cluster sampling if not enough $.
Choose area sampling.
Choose double sampling.
Choose convenience sampling.
All subgroups have equal number of elements?
Looking for information that only a few "experts" can provide?
Need responses of special interest minority groups?
Yes
No
Choose judgment sampling.
Choose quota sampling.
Choose proportionate stratified random sampling.
Choose disproportionate stratified random sampling

Precision and Confidence in Determining Sample Size

- Sample statistics should be reliable estimates and reflect the population parameters as closely as possible within an acceptable margin of error
- No sample statistic ($\bar{X}$) is going to be exactly the same as the population parameter (μ)
- Sample design is done to increase the probability that the sample statistics will be as close as possible to the population parameters

Precision

- **Precision – refers to how close our estimate is to the true population characteristic.**
- **Precision is a function of the range of variability in the sampling distribution of the sample mean**
- **Different samples give different sample means, called dispersion**
- **The smaller the dispersion or variability, the greater the probability that the sample mean will be closer the population mean**
- **This variability is called the standard error, calculated by,**

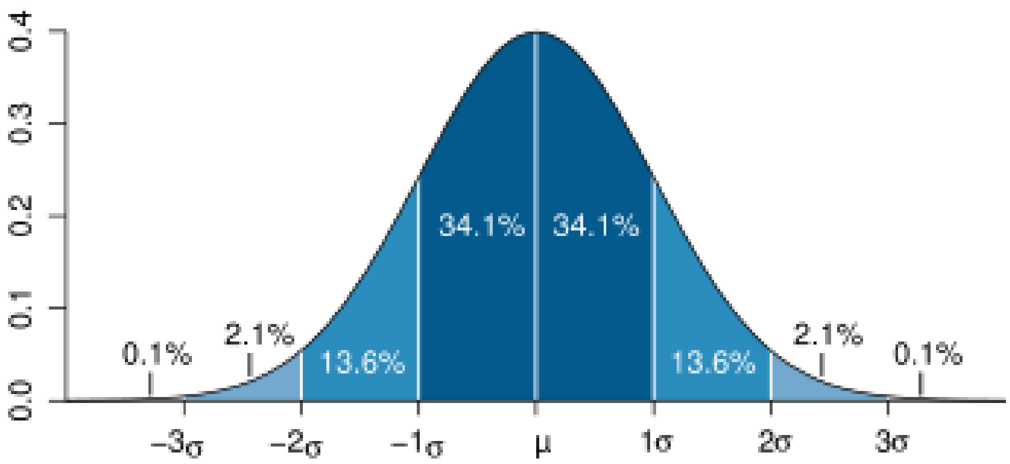
$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

S = the std. deviation of the sample

***n* = the sample size**

Confidence

- **Confidence – denotes how certain we are that our estimates will really hold true for the population**
- **Confidence reflects the level of certainty with which we can state that our estimates of the population parameters, based on our sample statistics, will hold true**



Confidence interval

- **Point estimation provides no measure of possible error.**
- **Therefore, an interval estimation is done to ensure a relatively accurate estimation of the population parameter**
- **Statistics that have the same distribution as the sampling distribution as the sampling distribution of the mean are used in this procedure, usually a z or a t static.**
- **The confidence interval around sample mean to estimate the range within which population mean would fall can be calculated using following formula where K is t statistic for the level of confidence desired.**

$$\mu = \bar{X} \pm K S_{\bar{X}}$$

Sample Data, Precision and Confidence in Estimation

- **To maintain the precision while increasing the confidence,**
- **To maintain the confidence level while increasing precision,**
- **To increase both the precision and the confidence,**
- **A larger sample size is needed. Therefore, sample size is a function of:**
  1. **the variability of population**
  2. **precision or accuracy needed**
  3. **confidence level desired**
  4. **type of sampling plan used**

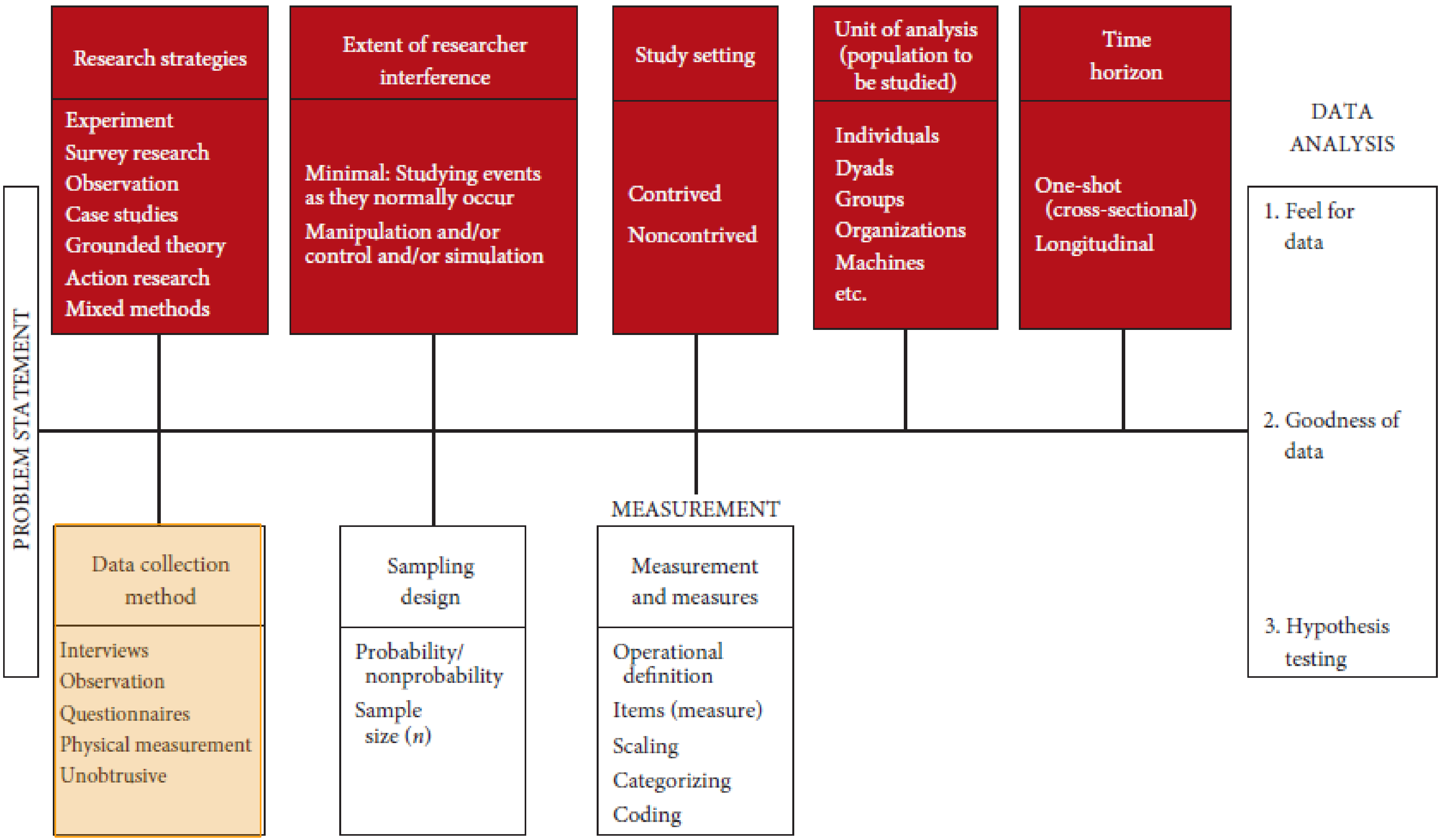
Thank You..!

Research Methods

Lecture 4-2: Data Collection

Agenda

- What is data collection
- Data collection techniques and tools
- Methods of data collection
- Observation
- Interviews
- Questionnaires
- Schedules


Research strategies
Experiment
Survey research
Observation
Case studies
Grounded theory
Action research
Mixed methods
Extent of researcher interference
Minimal: Studying events as they normally occur
Manipulation and/or control and/or simulation
Study setting
Contrived
Noncontrived
Unit of analysis (population to be studied)
Individuals
Dyads
Groups
Organizations
Machines
etc.
Time horizon
One-shot (cross-sectional)
Longitudinal
DATA ANALYSIS
1. Feel for data
2. Goodness of data
3. Hypothesis testing
PROBLEM STATEMENT
MEASUREMENT
Data collection method
Interviews
Observation
Questionnaires
Physical measurement
Unobtrusive
Sampling design
Probability/ nonprobability
Sample size (n)
Measurement and measures
Operational definition
Items (measure)
Scaling
Categorizing
Coding

What is data collection

- The process of gathering raw facts or evidence about a subject of interest in a systematic way, that enables one to answer stated research questions and evaluate outcomes
- The task of data collection begins after a research problem has been defined and research design chalked out
- Data collection tools are instruments that are used hand in hand with the data collection methods to collect/retrieve data

Data collection techniques and tools

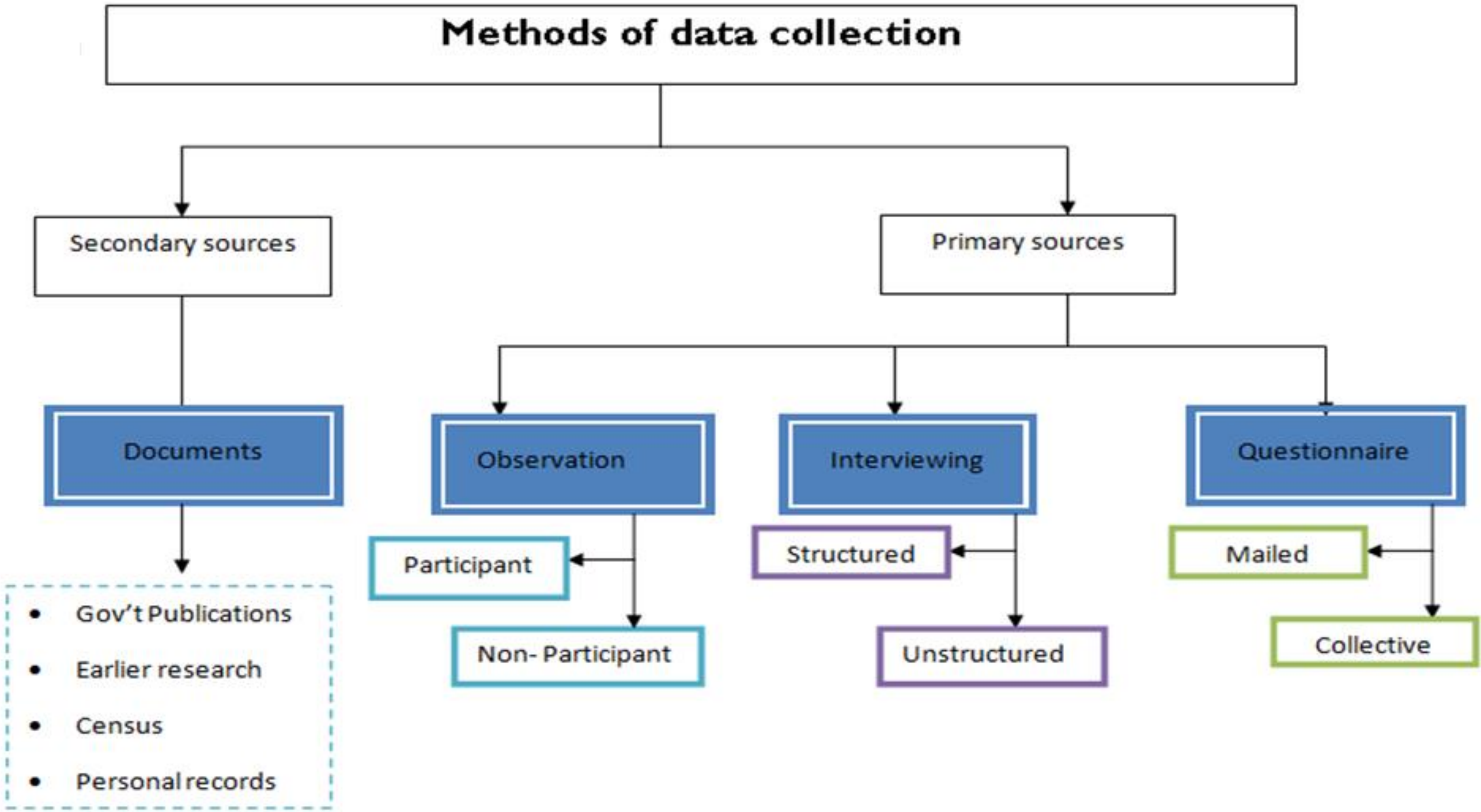
Techniques	Tools
• Administering written questionnaires.	• Questionnaire.
• Interviewing	• Interview guide, checklist, tape recorder.
• Observation	• Eyes and other senses, pen/paper, microscope.
• Using available information	• Check lists, data compilation forms.

Researcher must decide

- Which data to collect?
- How to collect the Data?
- Who will collect the Data?
- When to collect the Data?

- **Selection of a method for collecting information depends on**
  - Resources available
  - Credibility
  - Analysis and reporting
  - Resources
  - And the skill of the evaluator

Methods of data collection



Primary data collection

- Primary data may be collected through
  - Experiments
  - Surveys (sample surveys or census surveys)
  - Observation
  - Interviews
  - Questionnaires
  - Schedules

Observation method

- **Observation method is a method under which data from the field is collected with the help of observation by the observer or by personally going to the field Normality of Distributions**

- **P.V Young said that**

  "Observation may be defined as systematic viewing, coupled with consideration of seen phenomenon."

Steps for an effective observation

Determine what needs to be observed

Select participants Random/Selected

Conduct the observation
(venue, duration, recording materials, take photographs)

Compile data collected

Analyze and interpret data collected

Observation categories

1. Structured and unstructured observation

– **Structured:** When the observation is characterized by a careful definition of the units to be observed, the style of recording the observed information, standardized conditions of observation and the selection of related data of observation

– **Unstructured:** When it takes place without the characteristics of structured

2. Participant and non-participant observation

– **Participant:** When the observer is member of the group which he is observing then it is Participant Observation

– **Non-participant:** When observer is observing people without giving any information to them then it is Non-Participant Observation

Observation categories

3. **Uncontrolled and controlled observations**
   - **Uncontrolled:** When the observation takes place in natural condition i.e., uncontrolled observation. It is done to get spontaneous picture of life and persons
   - **Controlled:** When observation takes place according to pre arranged plans, with experimental procedure then it is controlled observation generally done in laboratory under controlled condition

Pros and cons of observations

Advantages	Disadvantages
• The observation technique can be stopped or begun at any time	• Interviewing selected subjects may provide more information, economically, than waiting for the spontaneous occurrence of the situation
• Produces Large quantities of data	• Expensive method
• Relative Inexpensive	• Limited information
	• Extensive Training is needed
	• All data obtained from observations are usable

Interview method

- The Interview Method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral- verbal responses.
- Interviewer asks questions (which are aimed to get information required for study) to respondent
- When adequate information obtained to understand the important factors operating in the situation, the researcher would stop the interviews
- Interviewers have to be thoroughly briefed about the research and trained in,
  - how to start an interview,
  - how to proceed with the questions,
  - how to motivate respondents to answer
  - what to look for in the answers,
  - How to close an Interview

Interview - questioning technique

- **Funneling:** Starting with broad open-ended questions and narrow down the focus when key issues relevant to the situation surfaced
- **Unbiased Questions:** Ask questions in a way that would ensure the least bias in the response
- **Clarifying Issues:** Restate or rephrase important information given by the respondent to get the intended response
- **Simple questions:** Help the respondent by asking the question in a simpler way
- **Taking Notes:** Researcher has to make written notes as the interviews are taking place

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Steps for an effective interview

Prepare interview schedule

↓

Select subjects/ key Respondent

↓

Conduct the interview

↓

Analyze and interpret data collected from the interview

Categorization of interviews

- **Structured Interviews:** A set of pre-decided questions is used
- **Unstructured Interviews:** Don't follow a system of pre-determined questions
- **Focused interviews:** Attention is focused on the given experience of the respondent and its possible effects
- **Group interviews**: a group of 6 to 8 individuals is interviewed
- **Individual Interviews:** Interviewer meets a single person and interviews
- **Qualitative and quantitative Interviews:** divided on the basis of subject matter i.e., whether qualitative or quantitative

Pros and cons of interviews

Advantages	Disadvantages
• Less expensive to administer mostly if the collective method is used	• An expensive Method
• Resistance may be overcome by a skilled interviewer	• Interviewer bias
• Personal information can be obtained	• Respondent bias
	• Time consuming

Questionnaire

- A questionnaire is a written list of questions, the answers to which are recorded by respondents
- In questionnaire respondents read the question, interpret what is expected and the write down the answers
- In the case of questionnaire, there is no one to explain the meaning of questions to respondents
- Therefore questions in a questionnaire must be clear and easy to understand

Administering a questionnaire

- ***A mailed questionnaire*** – one approach to collecting information is to send the questionnaire to prospective respondents by mail/email link (survey monkey). Obviously, this method presupposes that you have access to their address. One of the major problem with this approach is the low response rate.
- ***Collective administration*** – one of the best ways of administering questionnaires is to obtain a captive audience such as students in a class room, people attending a function, or people assembled in one place. This ensures a high response rate as you will find very few people refusing to participate in your study.

Cover letter of the questionnaire

- **It is essential that you write a cover letter to accompany your questionnaire. It should very briefly:**
  - Introduce you and the institution you are representing.
  - Describe in two or three lines the main objective of the study.
  - Convey any general instructions.
  - Assure respondents of the anonymity of the information provided by them.
  - Provide a contact number incase they have questions.
  - Give them a return address for the questionnaire and a deadline for its return.
  - Thank them for their participation.

Forms of questions

- **In a *closed-ended* question, the possible answers are set out in the questionnaire and the respondent is to choose the best alternative.**
  Types:- Dichotomous questions and Multiple Questions
  – *Please tick the type of research you would love to undertake.*
    i) Applied research ii) Basic Research iii) descriptive research
    iv) quantitative
- **In *open ended* questions, the possible responses are not given. The respondent writes down the answers in his/her words.**
  – *Please explain why you opted for a Bachelors degree in Information technology?*

Essentials of good questionnaire

- Should be short and simple
- Follow a sequence of questions from easy to difficult one
- Technical terms should be avoided
- Should provide adequate space for answers in questionnaire
- Directions regarding the filling of questionnaire should be given
  Physical Appearance – Quality of paper, Color
- Sequence must be clear
- Do not ask double barreled questions. A double barreled question is a question within a question._E.g. “How often and how much time do you spend in the computer lab”?
- Do not ask leading questions – these are questions in which, by their nature, contents, structure or wording leads a respondent to answer in a certain direction. E.g. “unemployment is increasing at a very high rate isn’t it” or “smoking is bad isn’t it?”

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Pros and cons of a questionnaire

Advantages	Disadvantages
• More information at greater depth can be obtained	• Limited application.
• Able to collect information from a larger sample	• Lack of opportunity to clarify issues.
• it offers greater anonymity.	• Low response rate
	• It is possible to consult others.
	• Spontaneous responses may not be possible mostly if it's a mailed questionnaire.

Schedules

- Very similar to Questionnaire method
- The main difference is that a schedule is filled by the enumerator who is specially appointed for the purpose.
- Enumerator goes to the respondents, asks them the questions from the Questionnaire in the order listed, and records the responses in the space provided
- Enumerator must be trained in administering the schedule

Questionnaire Vs. Schedule

Questionnaire	Schedule
• Generally send to through mail and no further assistance from sender	• Schedule is filled by the enumerator or research worker
• A cheaper method	• Costly requires field workers
• Non response is high	• Non response is low
• Incomplete and wrong information is more	• Depends on Honesty of the enumerator
• Depends on the quality of questionnaire	• Relatively more correct and complete

Thank You..!

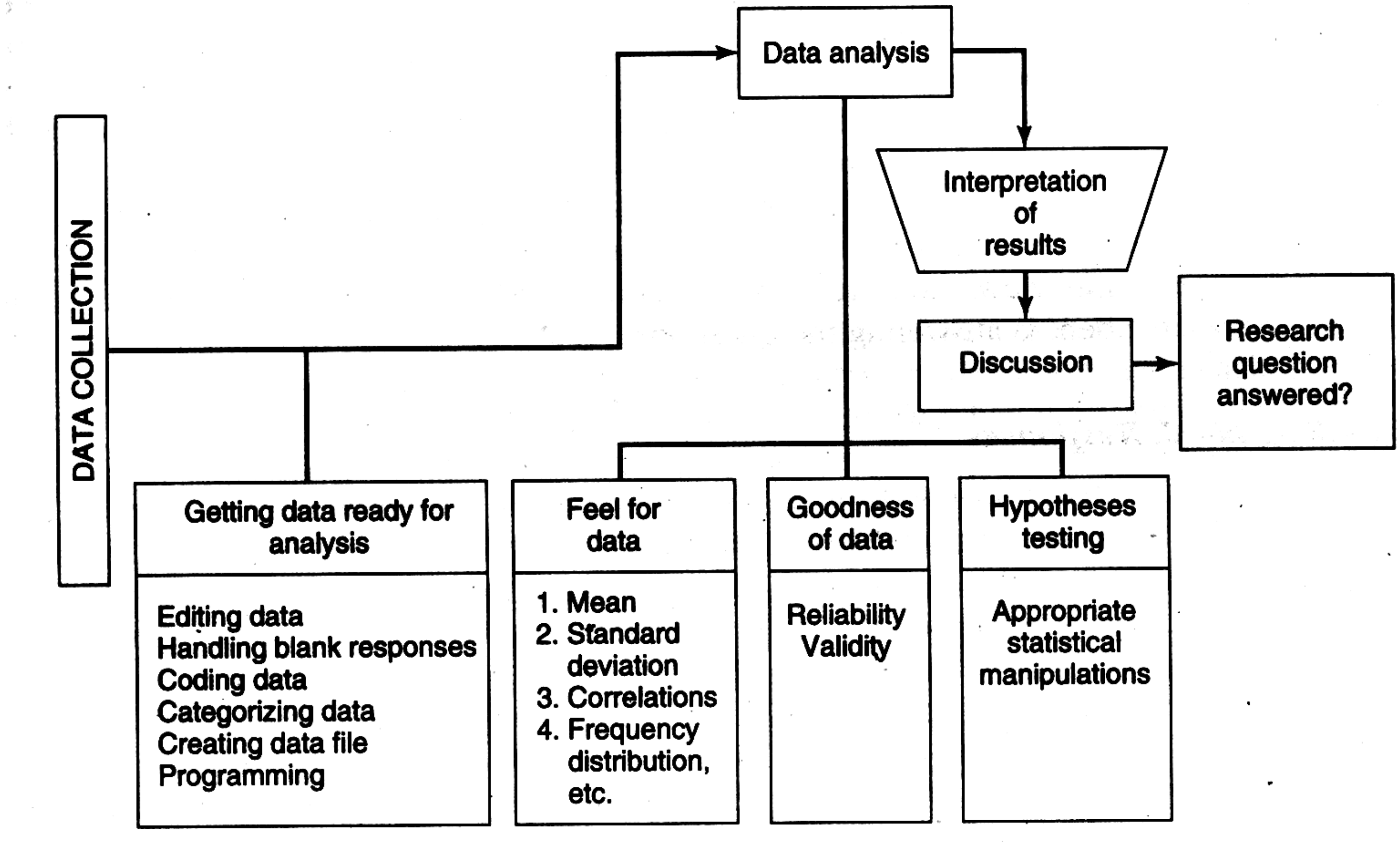
Research Methods

Lecture 5-1: Data Analysis

Agenda

- Getting data ready for analysis
  - Editing data
  - Handling missing data
  - Coding data
  - Categorizing data
  - Entering data
- Data analysis
- Introduction to statistics
- Statistical tests

Flow diagram of data analysis



Editing data

- Responses to open ended questions need to be edited
- Information that may be noted down by the interviewer, observer, or researcher in a hurry must be clearly deciphered
- Recommend to edit data on the very same day that the data are collected
  - Before forgetting
  - Enabling to contact the respondents for further information or clarification
- Editing should be identifiable through the use of a different colour pencil or ink so that the original information is still available in case of further doubts later
- Data have to be checked for incompleteness and inconsistencies
  - Check before make corrections if possible (otherwise bias may affect the goodness of data)

Handling missing data

- Reasons to have blank responses
  - Could not understand the problem
  - Did not know the answer
  - Was not willing to answer
  - Was simply indifferent to the need to respond the entire questionnaire
- If 25% of the questionnaire is not answered, better to throw it without using it for data analysis
- Important to mention number of returned but unused responses due to excessive missing data (final report submitted to the sponsor)

Handling blank responses

- An interval-scaled item with a mid-point (neutral value) would be to assign the midpoint in the scale as the response to that particular item
- Allow the computer to ignore the blank responses when analyzing (reduce sample size)
- Assign to the item the mean value of the responses of all other responded to that particular item
- Mean of the responses of this particular respondent to all other questions measuring this variable
- Random number within the scale
- Linear interpolation from adjacent point
- Linear trend in SPSS

Coding data

- Convenient to use scanner sheets to input data than keying them. When scanning is not possible due to some reason, it is better to use coding sheet first to transcribe data from the questionnaire and then key in the data (avoid flipping large questionnaires)
- Efficient when some thought is given to coding at the time of designing questionnaire
- Human errors can occur while coding. At least 10% of the coded questionnaires should therefore be checked for coding accuracy
- Use one of the rating scales

Rating scales

- Dichotomous scale
- Category scale
- Likert scale
- Numerical scale
- Semantic differential scale
- Itemized rating scale
- Fixed or constant sum rating scale
- Stapel scale
- Graphic rating scale
- Consensus scale

Categorizing data

- Set up a scheme for categorizing the variables such that the several items measuring a concept are all grouped together
- Responses to some of the negatively worded questions have also be reversed to make in same direction (eg: Time at work flies by quickly and my supervisor is not capable, working here is a drag)
- Questions measuring a concept is not contiguous (scattered) care has to be taken to include all the items without any omission or wrong inclusion

Entering data

- Scanning input or manually keying of data into the computer
- Any software can be used but spread sheet type may be preferred (row → case and column → variable)
- All missing values will appear with a dot in the cell

Objectives of data analysis

- **Getting feel for the data**
- **Testing the goodness of data**
- **Testing the hypothesis developed for research**

Getting feel for data

- Better to obtain
  - Frequency distributions for the demographic variables
  - Central tendency measures (mean, median, mode) and dispersion measures (standard deviation, range and variance, absolute deviation) on the other dependent and independent variables
  - An intercorrelation matrix of the respective variables, irrespective of whether or not the hypotheses are directly related to the analysis
- If an item has a little variation, researcher would suspect that particular question was not properly worded (not understand the intent of question). Otherwise it could be explained
- Graphical charts may be much easy
  - Histograms, bar charts
- Good to know how the dependent and independent variables in the study related to each other (inter-correlation matrix)
- If the correlation between two variables happens to be high (>0.75) check them for two different concepts

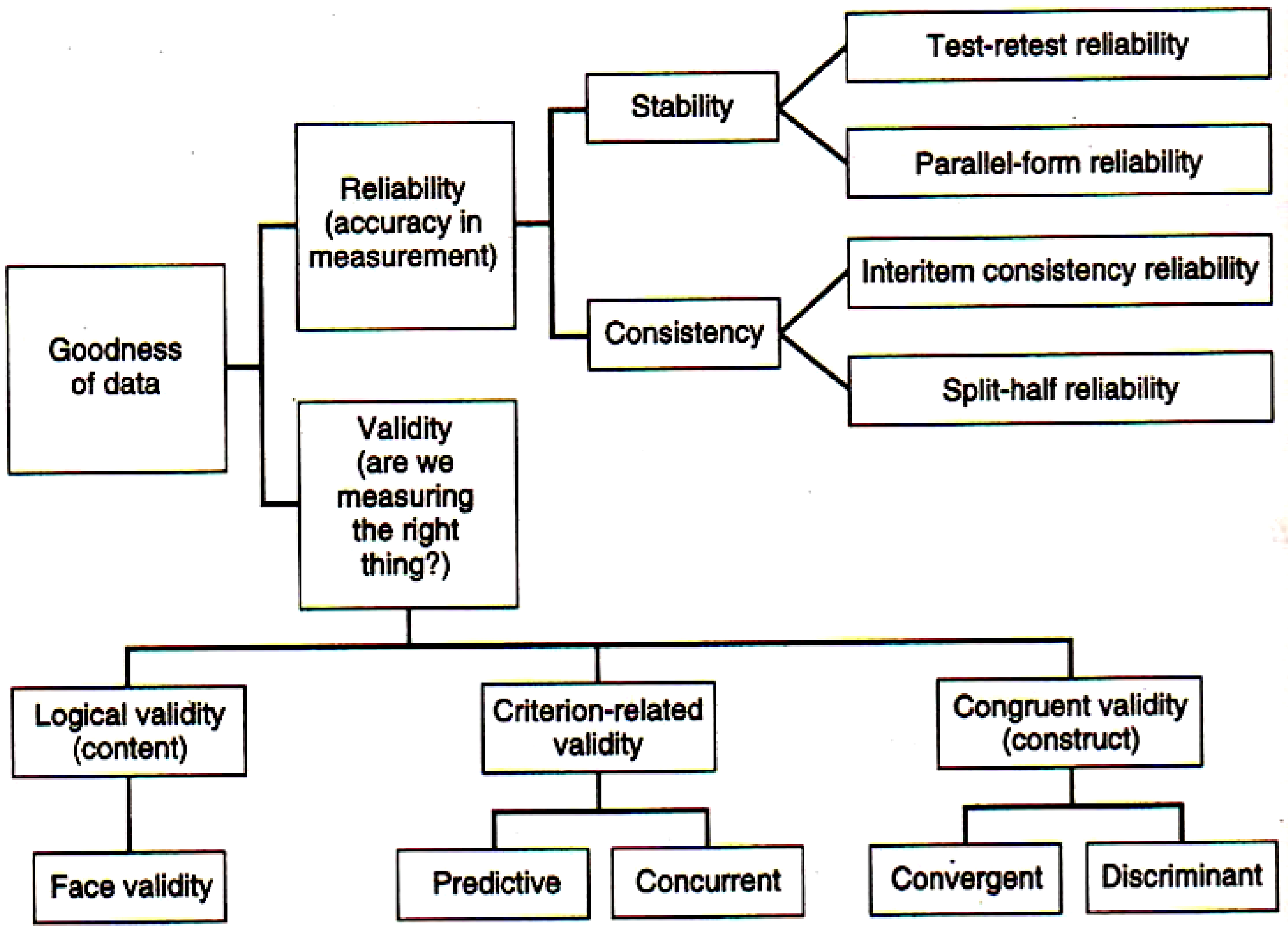
Testing goodness for data

- **Reliability and validity is tested**
- **Reliability**
  - Measure of consistency and stability
  - Consistency indicates how well the items measuring a concept hang together as a set
  - Cronbach's alpha is a reliability coefficient that indicates how well the items in a set are positively correlated to one another
    - **The closer Cronbach's alpha is to one, the higher the internal consistency reliability**
  - Another measure of consistency reliability used in specific situations: split half reliability coefficient which reflects the correlation between two halves of a set of items (vary depend on how the scale is split)
  - Split half reliability coefficient is obtain to test for consistency when more than one scale, dimension or factor is assessed.
  - Splitting should be based on predetermined logic
  - In almost every case, Cronbach's alpha is an adequate test for internal consistency reliability
  - The stability of a measure can be assessed through parallel form reliability and test-retest reliability

Validity

- Factorial validity → Factor analysis
- Results of factor analysis will confirm whether or not the theorized dimensions emerge (delineating the dimensions to operationalize the concept)
- Convergent validity can be established when there is high degree of correlation between two different sources responding to the same measure (eg: both supervisors and subordinates respond similarly)
- Discriminant validity can be established when two distinctly different concepts are not correlated to each other (courage and honesty, leadership and motivation) : Multi-trait multi-method matrix

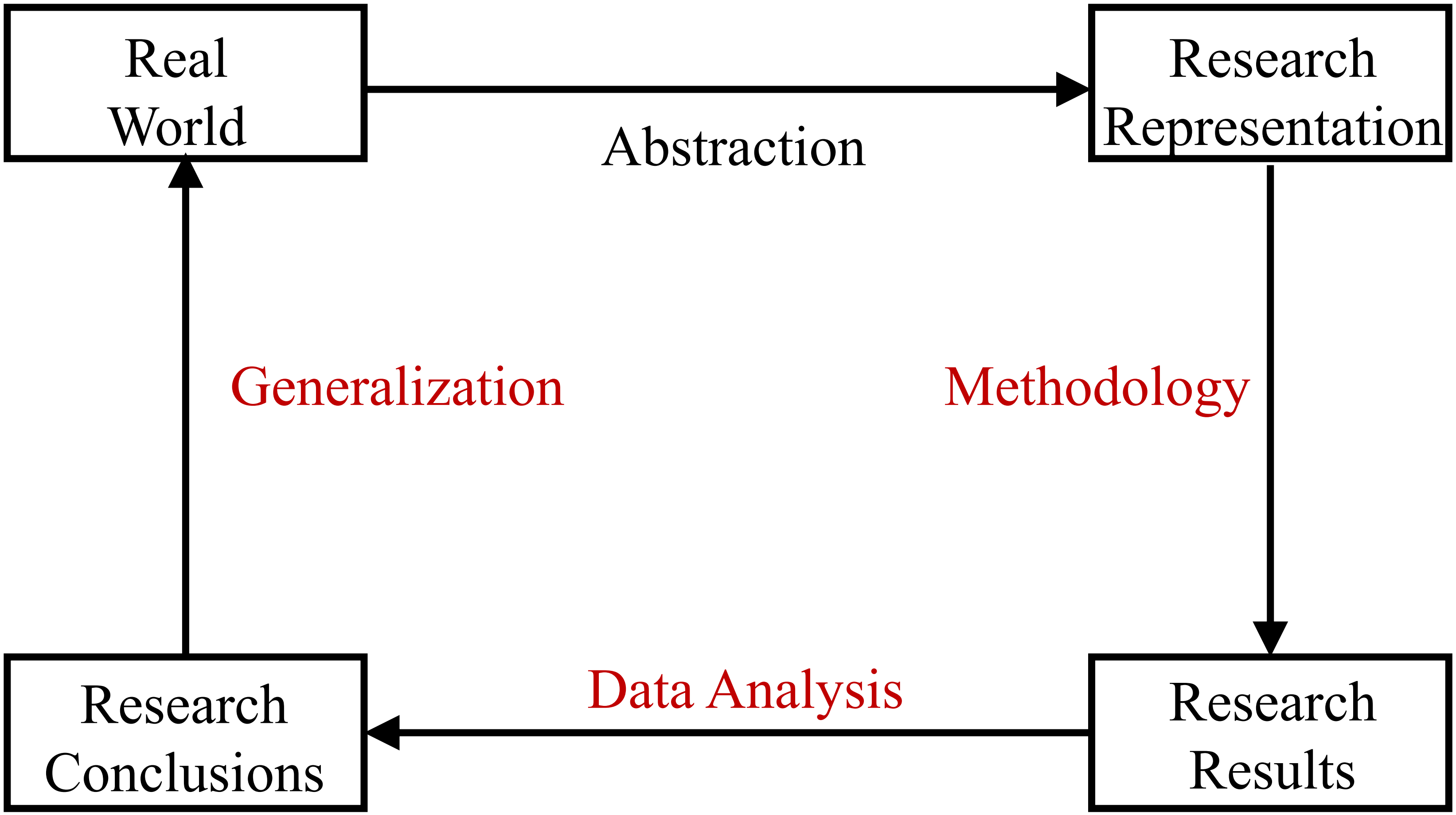
Forms of reliability and validity



Hypothesis testing

- Once the data are ready for analysis the researcher is ready to test the hypothesis already developed for the study
- Different tests that would be appropriate for different hypothesis and for data obtained on different scales
- In hypothesis testing it is essential to look at the followings
  - One tail or two tail test
    - To tail: equal null hypothesis and not equal alternative
    - One tail: less than null hypothesis and greater than alternative or vise versa
  - Sample size
    - N>25 normal distribution
    - N<25 t-distribution
  - Sample or population variance is known
    - Population: normal distribution
    - Sample: t-distribution
  - Generalization or sample comparison (check the underlying assumptions)

Statistics in research cycle



Introduction to Statistics

- **Statistics is a branch of mathematics dealing with data collection, organization, analysis, interpretation and presentation**
- **Data is a mixture of casual effects and causal effects. The statistics can filter out the causal effects from casual or random variation among data**
- **Statistics can be divided into the following categories**
  - Descriptive statistics: Frequency, central tendency and dispersion measures can be calculated.
  - Predictive statistics: Identification of clustering, regression among variables
  - Inferential statistics: Generalization from sample to population and this could be of two types namely estimations (point and interval) and hypothesis testing

Descriptive statistics-frequency

- Simply refer to number of times various subcategories of a certain phenomenon occurs
  - percentage and cumulative percentage can be calculated
- Information can graphically represented as a histogram or bar chart
- Examples
  - A marketing manager wants to know how many units of each brand of coffee are sold
  - Desire to obtain frequencies on a nominally scaled variable (grouped into non-overlapping subcategories)
  - In management research, frequencies are generally for nominal variables such as gender and educational level

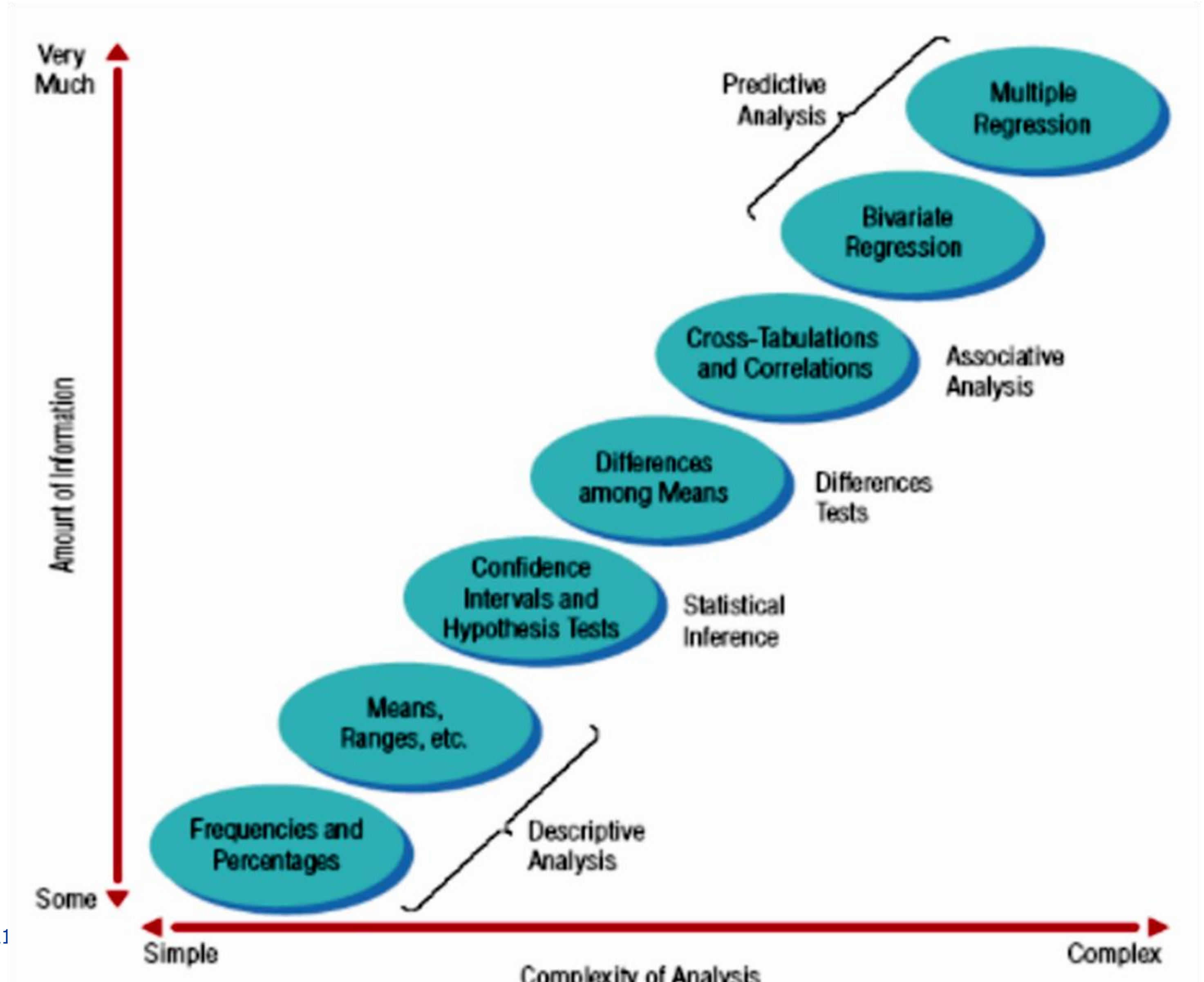
Descriptive statistics-central tendency

- **Mean:**
  - mean is the sum of the data set divided by the number of number of data
  - Mean value is most commonly referred to as the average.
  - Mean measures of central location and sensitive to outliers.
- **Median**
  - Median is the value of the point which has half the data smaller than that point and half the data larger than that point.
  - Good measure if the data set is skewed or having outliers
- **Mode**
  - Mode is the most frequently occurring phenomenon
  - It is not necessarily unique.
  - Mode is typically used in a qualitative fashion
  - It is the midpoint of the class interval of the histogram with the highest peak

Descriptive statistics-Dispersion

- Range:
  - Range is the largest value minus the smallest value in a data set.
  - This measure is based only on the lowest and highest extreme values in the sample.
  - The spread near the center of the data is not captured at all
- Variance
  - The variance is roughly the arithmetic average of the squared distance from the mean.
  - Squaring the distance from the mean has the effect of giving greater weight to values that are further from the mean.
  - Although the variance is intended to be an overall measure of spread, it can be greatly affected by the tail behavior.
- Standard deviation
  - Positive square root of the variance
  - Offers an index of the spread of a distribution or the variability in the data

Types of data analysis



Statistical tests

- **Based on the scale of variables the following descriptive analyses are meaningful**

OK to compute....	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean	No	No	Yes	Yes
ratio, or coefficient of variation	No	No	No	Yes

Selection of statistical tests

Variates			Criterion Variables: One – Nominal	One – Ordinal	One – Interval	Two or More – Nominal	Two or More – Ordinal	Two or More – Interval
Variates	One	Nominal	Chi-square test for inde-pendence Cochran Q test Fisher exact probability	Sign test Median test Mann-Whitney U test Kruskal-Wallis one-way analy-sis of variance	Analysis of variance			Multiple dis-criminant analysis
		Ordinal		Spearman's rank correlation Kendall's rank correlation	Analysis of variance with trend analysis			
	Two or More	Interval	Analysis of variance		Regression analysis	Analysis of variance		Multiple regression analysis
		Nominal		Friedman two-way analysis of variance	Analysis of variance (fac-torial design)			Analysis of variance
		Ordinal						
		Interval	Multiple discriminant analysis		Multiple regression analysis		Multiple discriminant analysis	Canonical correlation

dapted from R. L. Baker & R. E. Schultz (Eds.) *Instructional product research*. New York: Van Nostrand Co., 1972, p. 110

Test	When Used	Function
Chi-square	With **nominal** data for one sample or two or more independent samples.	Tests for independence of variables.
Cochran Q	With more than two related samples measured on **nominal** scale.	Helps when data fall into two natural categories.
Fisher exact probability	With two independent samples measured on **nominal** scale.	More useful than χ^2 when expected frequencies are small.
Sign test	With two related samples measured on **ordinal** scale.	A good test for ranked data.
Median test	With one sample, to see if randomly drawn measurements are from a population with a specified median.	In a symmetric distribution, the mean and median will be the same.
Mann-Whitney U test	With two independent samples on **ordinal** data.	Analogue of the two independent sample *t*-tests.
Kruskal-Wallis one-way ANOVA	With more than two independent samples on an **ordinal** scale.	An alternative to one-way ANOVA where normality of distributions cannot be assumed.
Friedman two-way ANOVA	With more than two related samples on **ordinal** data.	A good alternative to two-way ANOVA where normality cannot be assumed.
Kolmogorov-Smirnov	With one sample or two independent samples measured on an **ordinal** scale.	Is a more powerful test than χ^2 or Mann-Whitney U.

Software packages for Data analysis

- **Popular Commercial Packages**
  - SPSS
  - SAS
  - Minitab
  - Matlab Statistical toolbox
  - PASS (power and sample size software)
  - Excel with statistical add ons
- **Popular Open Source Packages**
  - PSPP (alternative to IBM SPSS)
  - DAP (free replacement for SAS)
  - R (Programmable with S language)
  - ADaMSoft (statistical software with data mining algorithms and methods for data management)

Thank You..!

Research Methods

Session: Qualitative Research for MSc Students

By
Prof. Samantha Thelijjagoda
Dean-SLIIT Business School

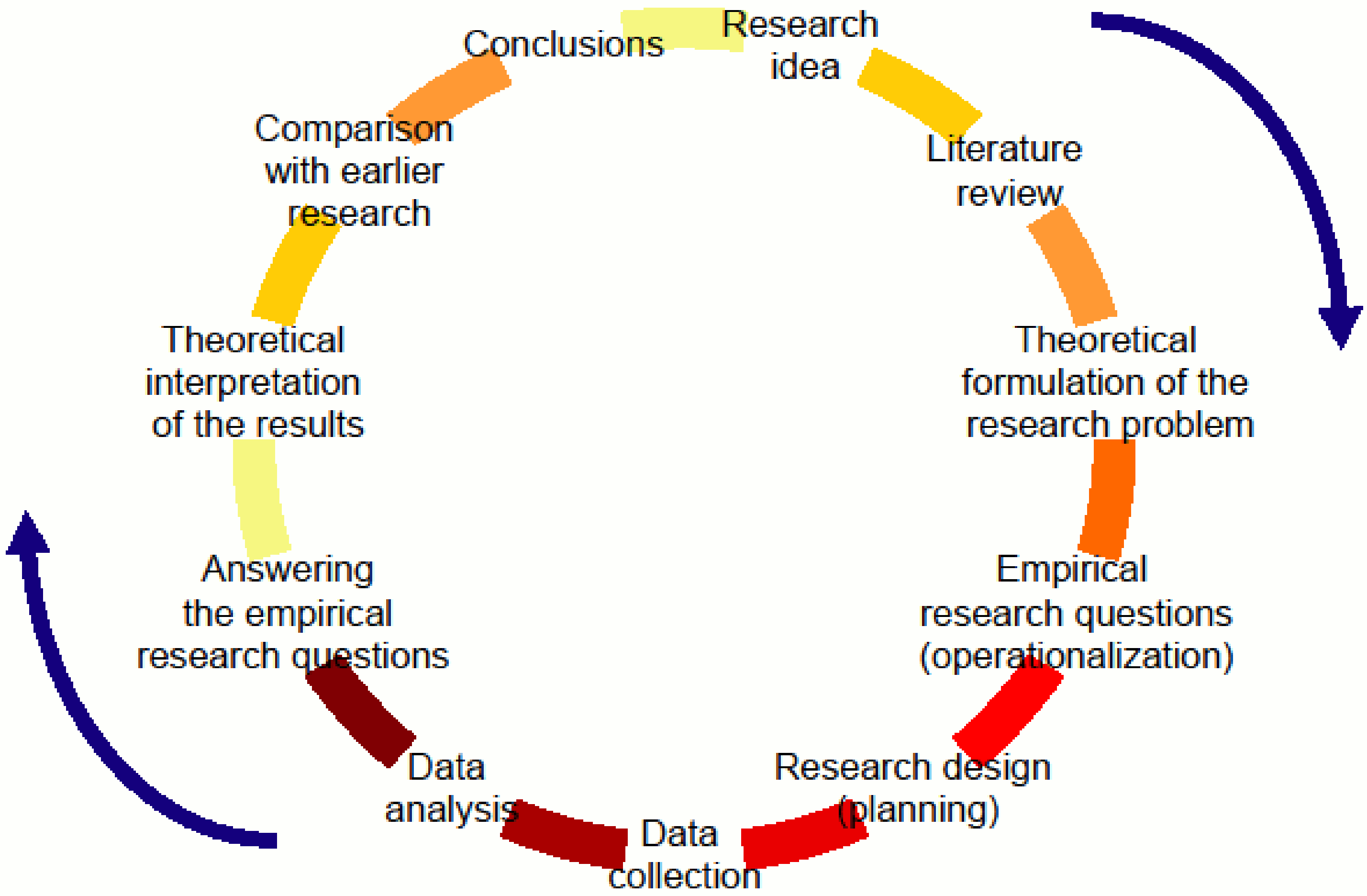
Outline

Learning Outcomes
Research Process
Qualitative research- Definition
Qualitative Research Methods
Collecting Evidences
Data Analysis in QR
Techniques of Analysis

Learning Outcomes

To understand the nature and approaches of QR
To recognize how to design a qualitative research
To understand how to analyze and present
-the results of qualitative data
To identify the challenges of doing qualitative research

The research process



What we do in Qualitative Research?

- We rely on data collected from interviews, observations, and content analysis of newspapers, books, videos, case records, and other already developed documents.

- We may try to develop new theories based on what happens in specific situations.

What we do in Qualitative Research?

Quantitative researchers apply general theories to revolve problems

Qualitative researchers:

- examine what happens in specific situations and
- try to develop new theories based on that situation.

What we do in Qualitative Research?

- Answer research questions rather than test a hypothesis.
- Examine the perceptions, actions, and feelings of participants.
- Obtained detailed information from interviews, content analysis, or observations.

Definition: Qualitative Research

In the handbook of qualitative research Denzin and Lincoln (2005) describe qualitative research as involving "… *an interpretive naturalistic approach to the world. This means that qualitative researchers study things in their natural settings, attempting to make sense of or interpret phenomena in terms of the meanings people bring to them."* (p. 3)

Qualitative research is a type of social science research that collects and works with non-numerical data and that seeks to interpret meaning from these data that help us understand social life through the study of targeted populations or places.

Qualitative Research Methods

Action research

Ethnography

Grounded theory

Case study method

Content analysis

Field research

Phenomenology

Historical Method

Action Research

Aims to contribute both the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework.

Is a disciplined process of inquiry conducted by and for those taking the action. The primary reason is to assist the “actor” in improving and/or refining his or her actions.

Example:

-After school programs and their effects on academic achievement

-Music and its impact on student's writing productivity & attitude toward writing.

Ethnographic Research

- Ethnography is the study of social interactions, behaviors, and perceptions that occur within groups, teams, organizations, and communities.
- Comes from the discipline of social and cultural anthropology
- Ethnographer is required to spend a significant amount of time in the field.
- Ethnographers immerse themselves in the lives of the people they study and seek to place the phenomena studied in their social and cultural context.

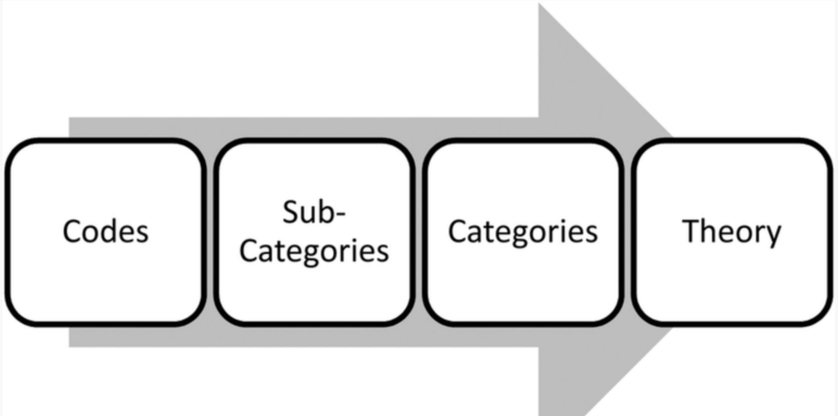
Grounded theory

Seeks to develop theory that is grounded in data systematically gathered and analysed.

An inductive, theory discovery methodology.

There is continuous interplay between data collection and analysis.

....to read a textual database and "discover" or label variables (called categories, concepts and properties) and their interrelationships.



Case Study

- Case study "Investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and multiple sources of evidence are used" (Yin 2003:13-14).

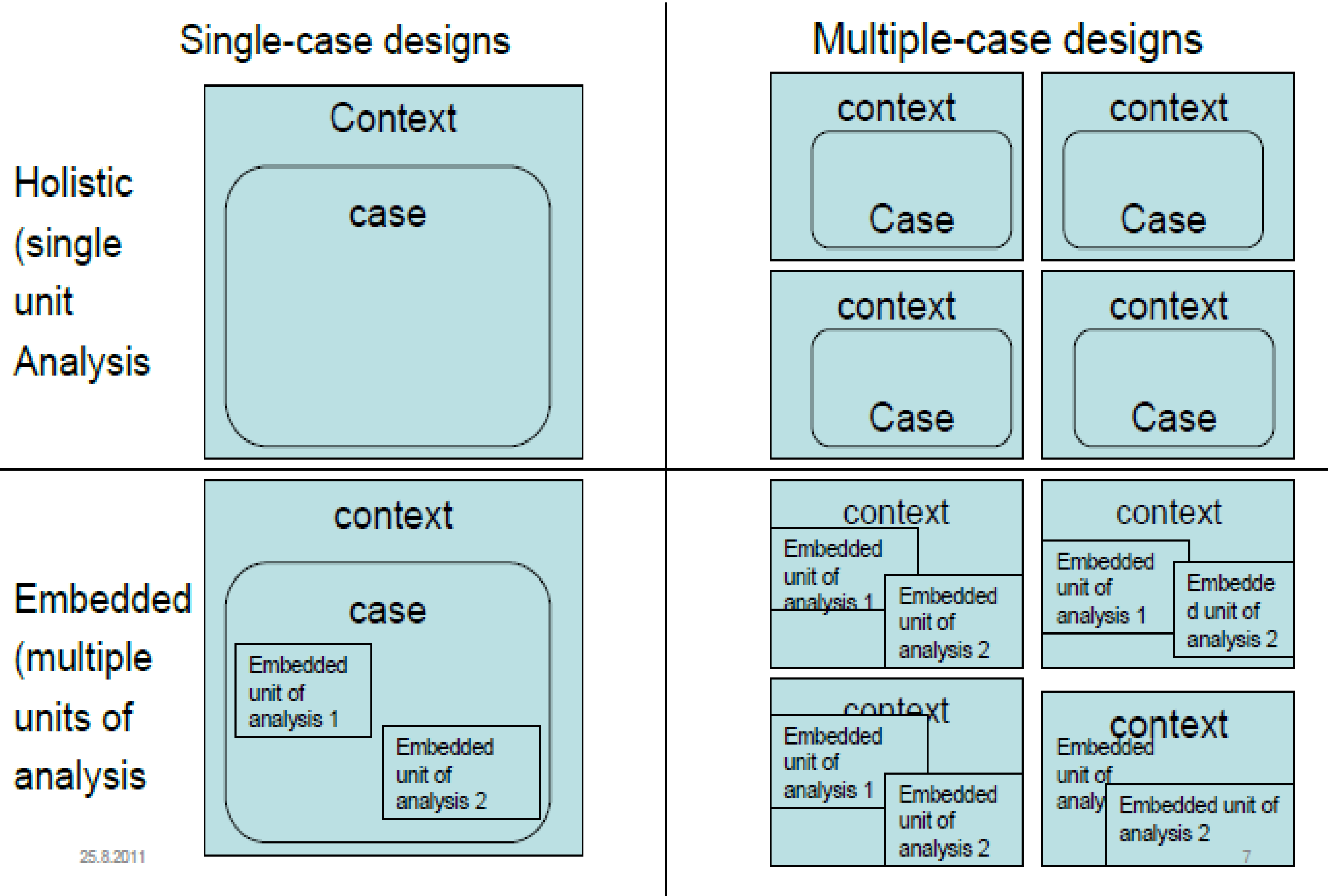
- Case study is an ideal methodology when a holistic, in-depth investigation is needed

  (Feagin, Orum, & Sjoberg, 1991).

Holistic and Embedded Case Studies

Single case - and multiple case studies can further be classified as holistic or embedded. In an embedded case study, the case is split in multiple units of analysis, while a holistic case study has one unit of analysis for each case.

Holistic and embedded case studies



Content Analysis

- Content analysis is used to analyse the presence, meanings and relationships of words and concepts in a text

- then make inferences about the messages within the texts, the writer(s), the audience, and even the culture and time of which these are a part

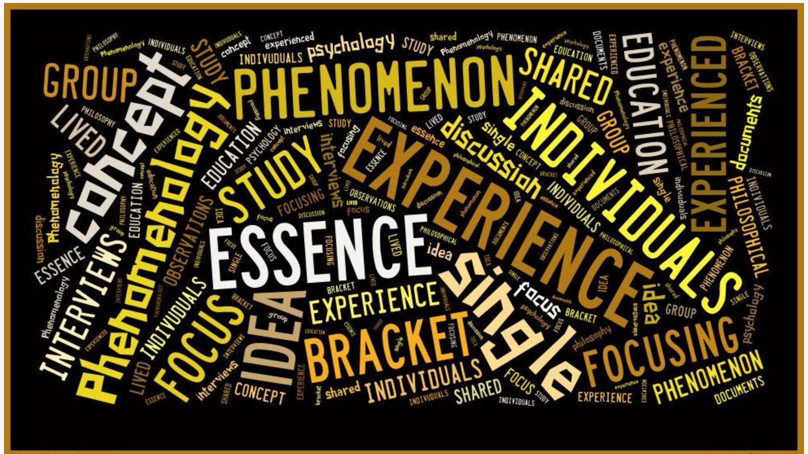
Field Research

- Field researcher goes directly to the social phenomenon under study and observes it as completely as possible.
- The natural environment is the priority of the field researcher. There are no implemented controls or experimental conditions to speak of.
- Such methodologies are especially useful in observing social phenomena over time.

Phenomenology

A phenomenological study describes the meaning of several individuals of their lived experience.

- "The study of the lived experiences of persons"
- Experience is a conscious process
- The "development of [interpretations] of the essences of these experiences



Historical Method

Systematic collection and objective evaluation of data related to past occurrences in order to explain causes, effects, or trends of these events in the present and anticipate future events.

A Comparison

	Qualitative	Quantitative
Type of Question	Why, How	What, Who
Sample Size	Small	Large
Information per Respondent	Much	Vary
Tools	Analytical	Statistical
Type of Research	Exploratory /Explanatory	Descriptive or Causal

Collecting Evidence

- There are six forms of evidence:
  - Documentation
  - Archival records
  - Interviews
  - Focus Group Discussion
  - Direct observations
  - Participant observations
  - Physical Artifacts
- Principles:
  - Using multiple, not single, sources of evidence
  - Maintaining a chain of evidence

Documents, Archives & Artifacts

- Current organisational documents
  - Policies, practices, templates
- Historical archive records
  - Past practices, reports on past projects
- These two forms of data are useful for corroborating evidence from other sources
- Artifacts include physical objects like certificates, awards, etc.

Interviews

- Case study interviews are usually open-ended
- Focused interviews can still be open-ended but involve following specific questions derived from the case study protocol
- More structured questions, along the lines of a formal survey.
- A structured interview would involve the sampling procedures and the instruments used in regular surveys, and it would subsequently be analyzed in a similar manner.

Focus Group Discussion

Collect information from groups of people rather than a series of individuals

- FGD can be used when
  - Resources are limited
  - To identify a number of individuals who share a common factor
  - It is desirable to collect the views of several people within the population sub group
  - Group interaction among participants has the potential for greater insights to be developed

Direct Observations

- By making a field visit to the "site", the investigator is creating the opportunity for direct observations.
- To increase the reliability of observational evidence, a common procedure is to have more than one observer making an observation, whether formally or casually.

Participant Observations

- The researcher may play a variety of roles within a situation and may actually participate in the events being studied.
- The researcher (e.g. if an insider) may be able to gain access to events or groups that are otherwise inaccessible to scientific investigation.
- Researchers have to be aware of potential bias. If they work as an insider, they cannot maintain outside independence.
- They could become a supporter or defender.

Data analysis in Qualitative Research

Consists of three concurrent flows of action:

- data reduction
- data display
- conclusions and verification

These flows are present in parallel during and after the collection of data (Miles and Huberman, 1994).

- **Data reduction** refers to the process of selecting, focusing, simplifying, abstracting and transforming the collected data. It needs to be reduced in order to make the data more readily accessible and understandable (Berg, 2004; Kvale, 1996).
- **Data display** is intended to organize the collected data in such a way that it permits conclusion drawing (Miles and Huberman, 1994; Berg, 2004).
- **Data analysis** process is conclusion drawing and verification. During the collection of data, there should not be made any definitive conclusions, and these preliminary conclusions should be verified during the process (Miles and Huberman, 1994).

Computerized Qualitative Data Analysis

atlas.ti

QUALITATIVE DATA ANALYSIS

NVIVO

QUALITATIVE DATA ANALYSIS

BY

Professor Samantha Thelijjagoda
Dean-SLIIT Business School
Sri Lanka Institute of Information Technology

Material Ownership:
Professor (Mrs) Felice D. Billups
University of Wales, UK

GETTING STARTED?

- Have you just conducted a qualitative study involving…
- Interviews
- Focus Groups
- Observations
- Document or artifact analysis
- Journal notes or reflections?

WHAT TO DO WITH ALL THIS DATA?

- Just as there are numerous statistical tests to run for quantitative data, there are just as many options for qualitative data analysis…

OVERVIEW

- This session is designed to provide a step-by-step guide for beginning qualitative researchers…who want to know how to apply the appropriate strategies for data analysis, interpretation, and reporting.

LIKE CLEANING A CLOSET ???

- Think of managing your qualitative analysis process like cleaning your closets – the same basic steps apply!

It's the same process…

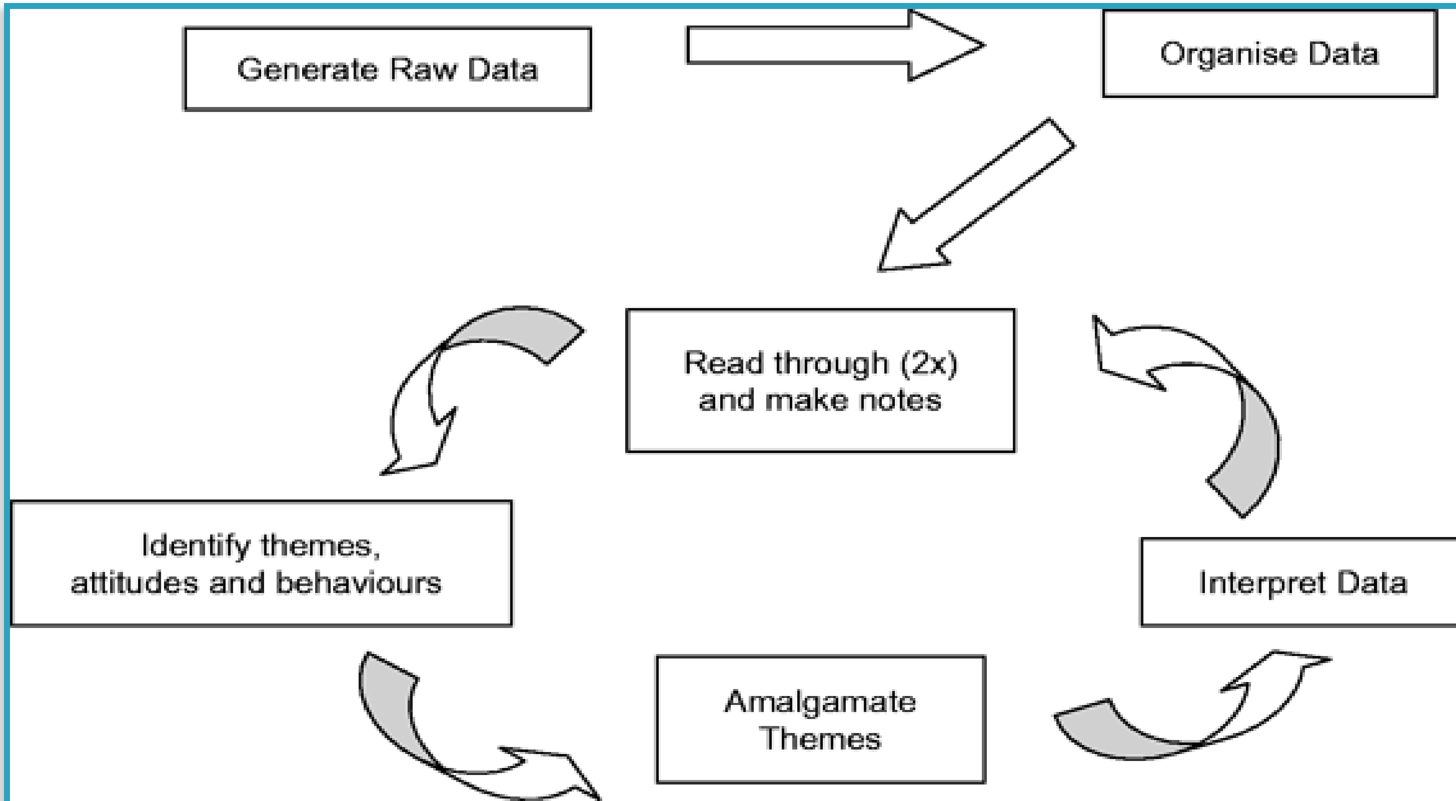
- 1. Take everything out of the closet
- 2. Sort everything out – save or throw ?
- 3. Look at what you have left and organize into sub-groupings (chunking)
- 4. Organize sub-groups into clusters of similar things that belong together (clusters, codes)
- 5. As you put things back, how would you group them to maximize functionality? How do the groups make it work together? (interpretation, presentation)

FOUR BASIC STEPS

- All qualitative data analysis involves the same four essential steps:

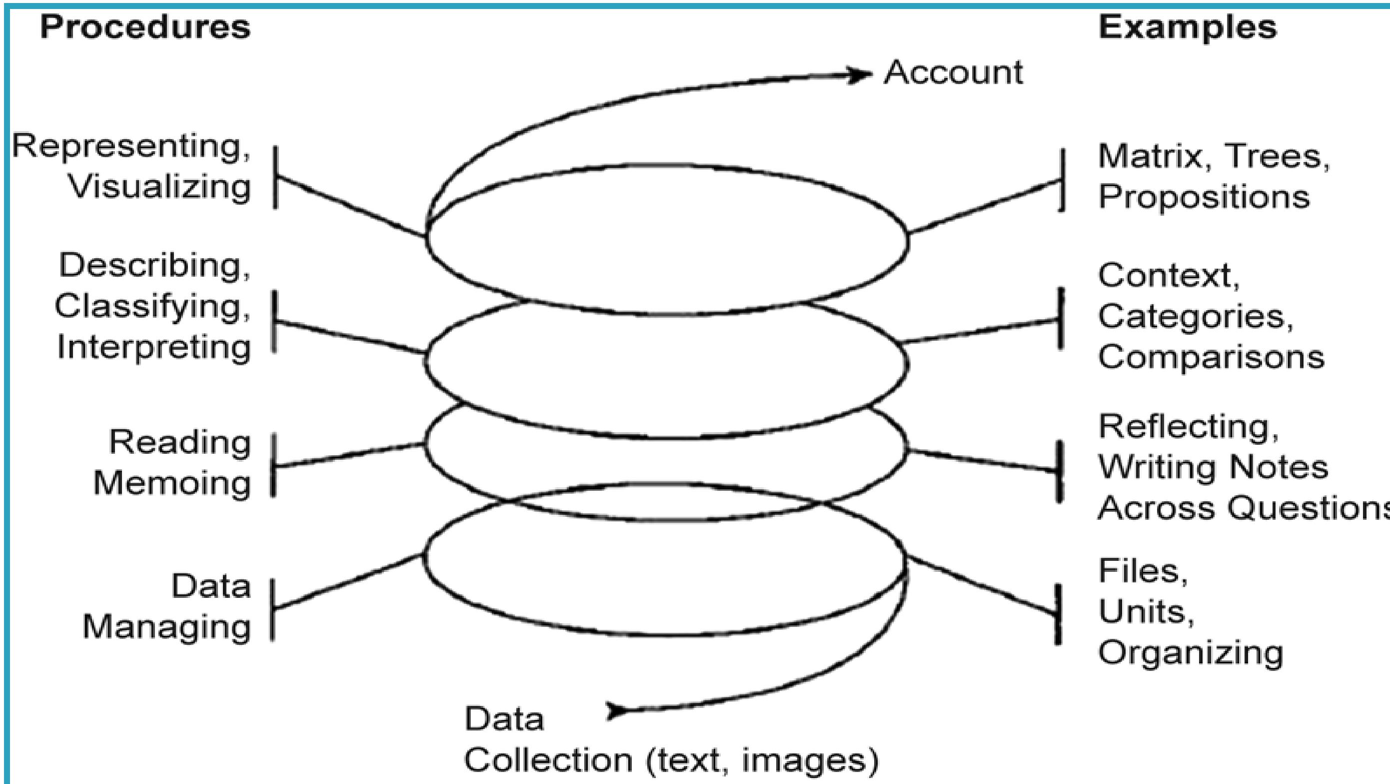
- 1. Raw data management– ‘data cleaning’
- 2. Data reduction, I, II – ‘chunking’, ‘coding’
- 3. Data interpretation – ‘coding’, ‘clustering’
- 4. Data representation – ‘telling the story’, ‘making sense of the data for others’

DATA ANALYSIS SPIRAL #1



Source: Adapted from Creswell (2009)

DATA ANALYSIS SPIRAL #2



Source: Creswell (2007)

Step 1: Raw Data Management

- What is raw data management?
  - The process of preparing and organizing raw data into meaningful units of analysis:
    - Text or audio data transformed into transcripts
    - Image data transformed into videos, photos, charts

As you review your data, you find that some of it is not usable or relevant to your study…

Raw Data Sample

- I always wanted to get my doctorate but I never felt I had the time; then I reached a point in my career where I saw that without the credentials, I would never advance to the types of positions I aspired to..but I doubted I could do the work. I wasn't sure I could go back to school after so much time. And did I have the time, with working and a family? These were the things I struggled with as I looked for the right program.
- Um, ..finally starting the program with others like me, it felt surreal. Once you switch gears from being an established administrator at a college to being a doc student, you realize you lose control over your life. You are not in charge in that classroom, like you are in your office. But also, once you say you are a doc student, people look at you differently. And people at work began to take me more seriously, ask for my opinion as if I now possessed special knowledge because I was going for the doctorate. It was the same information I had shared previously but somehow it had a special quality? Its like magic!
- I can't think of a particular example right now…

- Are some portions of this transcript unusable or irrelevant? (purple)

Transcript of Interview Data

Raw Data Overview

Step II: Data Reduction I

- Get a sense of the data holistically, read several times (immersion)
- Classify and categorize repeatedly, allowing for deeper immersion
- Write notes in the margins (memoing)
- Preliminary classification schemes emerge, categorize raw data into groupings (chunking)

Winnowing

- Develop an initial sense of usable data and the general categories you will create
- Preliminary set of codes developed, cluster raw data into units that share similar meanings or qualities
- Create initial code list or master code book

Chunks-Clusters Sample

- I always wanted to get my doctorate but I never felt I had the time; then I reached a point in my career where I saw that without the credentials, I would never advance to the types of positions I aspired to..but I doubted I could do the work. I wasn't sure I could go back to school after so much time. And did I have the time, with working and a family? These were the things I struggled with as I looked for the right program.
- -finally starting the program with others like me, it felt surreal. Once you switch gears from being an established administrator at a college to being a doc student, you realize you lose control over your life. You are not in charge in that classroom, like you are in your office. But also, once you say you are a doc student, people look at you differently. And people at work began to take me more seriously, ask for my opinion as if I now possessed special knowledge because I was going for the doctorate. It was the same information I had shared previously but somehow it had a special quality? Its like magic!

Transcript of Interview Data

- Which sections of data are broadly similar? (red for credentials , blue for personal struggles, green for shift in identity)
- Which 'chunks' can be clustered together to relate to a broad coding scheme?

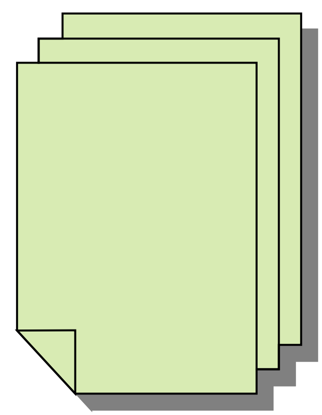
Chunking? Clusters?

Step II: Data Reduction II

- The process of reducing data from chunks into clusters and codes to make meaning of that data:
  - Chunks of data that are similar begin to lead to initial clusters and coding
    - Clusters – assigning chunks of similarly labeled data into clusters and assigning preliminary codes
    - Codes – refining, developing code books, labeling codes, creating codes through 2-3 cycles

Coding Process

- Initial coding may include as many as 30 categories
- Reduce codes once, probably twice
- Reduce again to and refine to codes that are mutually exclusive and include all raw data that was identified as usable



A Priori or In Vivo Codes

- A Priori (pre-set codes)
  - Codes derived from literature, theoretical frames
- In Vivo (labeling a section of data) (inductive or grounded)
  - Codes derived from the data by using code names drawn from participant quotes or interpretation of the data
    - "Its like magic" is a phrase that could form the basis for a code category

Coding Levels

- Descriptive to Interpretative to Pattern Coding
  - Moves from summary to meaning to explanation
- OR
- Open to Axial to Selective Coding
  - Moves from initial theory to developing relationships between codes for emerging theory
- OR
- First cycle to second cycle coding
  - Moving from describing the data units to inferring meaning

Coding Sample

- I always wanted to get my doctorate but I never felt I had the time; then I reached a point in my career where I saw that without the credentials, I would never advance to the types of positions I aspired to..but I doubted I could do the work. I wasn't sure I could go back to school after so much time. And did I have the time, with working and a family? These were the things I struggled with as I looked for the right program.
- -finally starting the program with others like me, it felt surreal. Once you switch gears from being an established administrator at a college to being a doc student, you realize you lose control over your life. You are not in charge in that classroom, like you are in your office. But also, once you say you are a doc student, people look at you differently. And people at work began to take me more seriously, ask for my opinion as if I now possessed special knowledge because I was going for the doctorate. It was the same information I had shared previously but somehow it had a special quality? Its like magic!

- Chunking to coding:
- Red for credentials – codes include career goals CG, career advancement CA
- Blue for personal struggles– codes include self-doubt SD, time management TM
- Green for shift in identity – codes include student role SR, identity at work IW, shift in control SC

Transcript of Interview Data

Chunking? Clusters? Coding?

Coding Levels (*revisiting*)

- Descriptive to Interpretative to Pattern Coding
  - Moves from summary to meaning to explanation
- **OR**
- Open to Axial to Selective Coding
  - Moves from initial theory to developing relationships between codes for emerging theory
- **OR**
- First cycle to second cycle coding
  - Moving from describing the data units to inferring meaning

Coding Progression

- <u>Descriptive   Interpretative</u>
- Credentials   CG,CA
  - need for career advancement, goals
- Personal   PSD,PG,PWL
  - Self-doubt
  - Personal growth
  - Work-life balance
- Identity   IS, ISR, ISC
  - identity shifting
  - student role
  - shift in control

- <u>Pattern</u>
- CR – needing a doctorate to advance professionally and to meet personal goals for achievement
- PG – personal struggles evolve to address self-doubt about abilities, trying to achieve things before time runs out, balancing responsibilities with family, self, work
- IS – managing the shift from student to graduate, from candidate to doctor, from non-expert to expert in work settings, from losing control to re-gaining control at home and work

Descriptive to Interpretative

Pattern – Inductive meaning

Step III: Data Interpretation & Themes

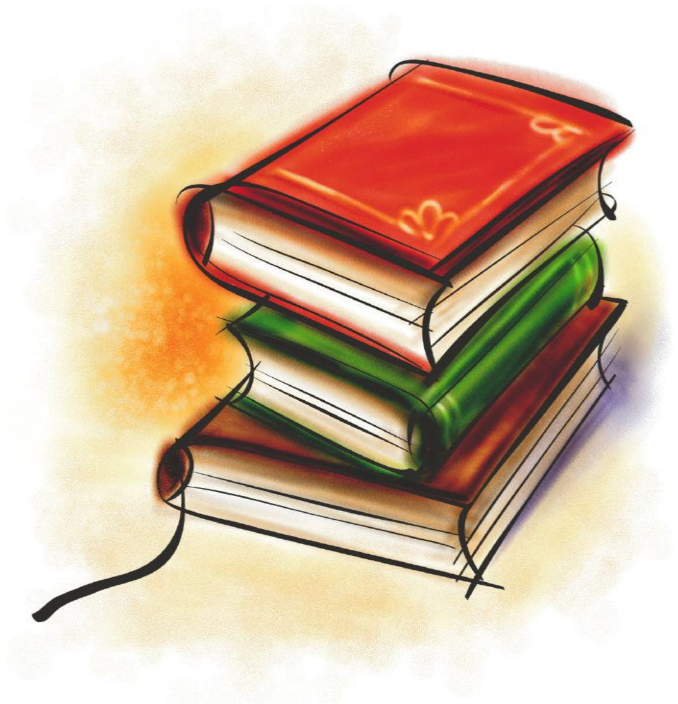
- ‘Chunks’ of related data that have similar meaning are coded in several cycles
- Once coded, those ‘chunks’ become clustered in similar theme categories
- Create meaning for those clusters with labels
- Themes emerge from those clusters
- Interpret themes to answer research questions

Step IV: Data Representation

- Interpretation or analysis of qualitative data simultaneously occurs
- Researchers interpret the data as they read and re-read the data, categorize and code the data and inductively develop a thematic analysis
- Themes become the story or the narrative

Data Representation Types

- Telling the story with the data
  - Storytelling, Narrative
  - Chronological
  - Flashback
  - Critical Incidents
  - Theater
  - Thematic
  - Visual representation
  - Figures, tables, charts



Computer Software

- ATLAS.ti, HyperRESEARCH, Nvivo, MaxQDA, NUD*IST
- Software packages either assist with theory-building or with concept mapping
- Data-voice recognition software converts audio into text, such as Dragon

References

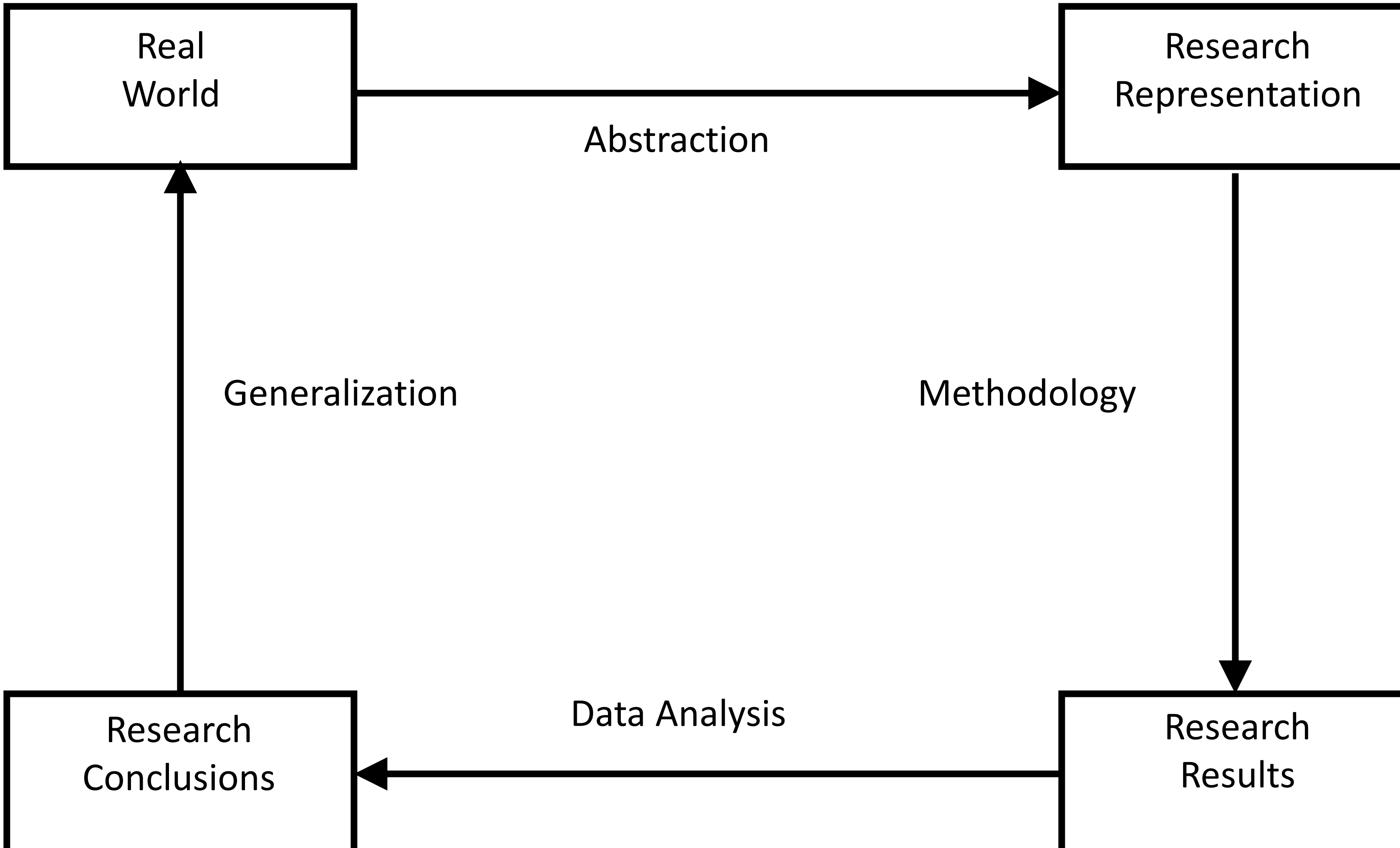
Grbich, C. (2007). Qualitative data analysis: An introduction. London, UK: Sage.

Miles, M. B., & Huberman, A. M. (2013). Qualitative data analysis: An expanded sourcebook. (3rd ed.). Los Angeles, CA: Sage.

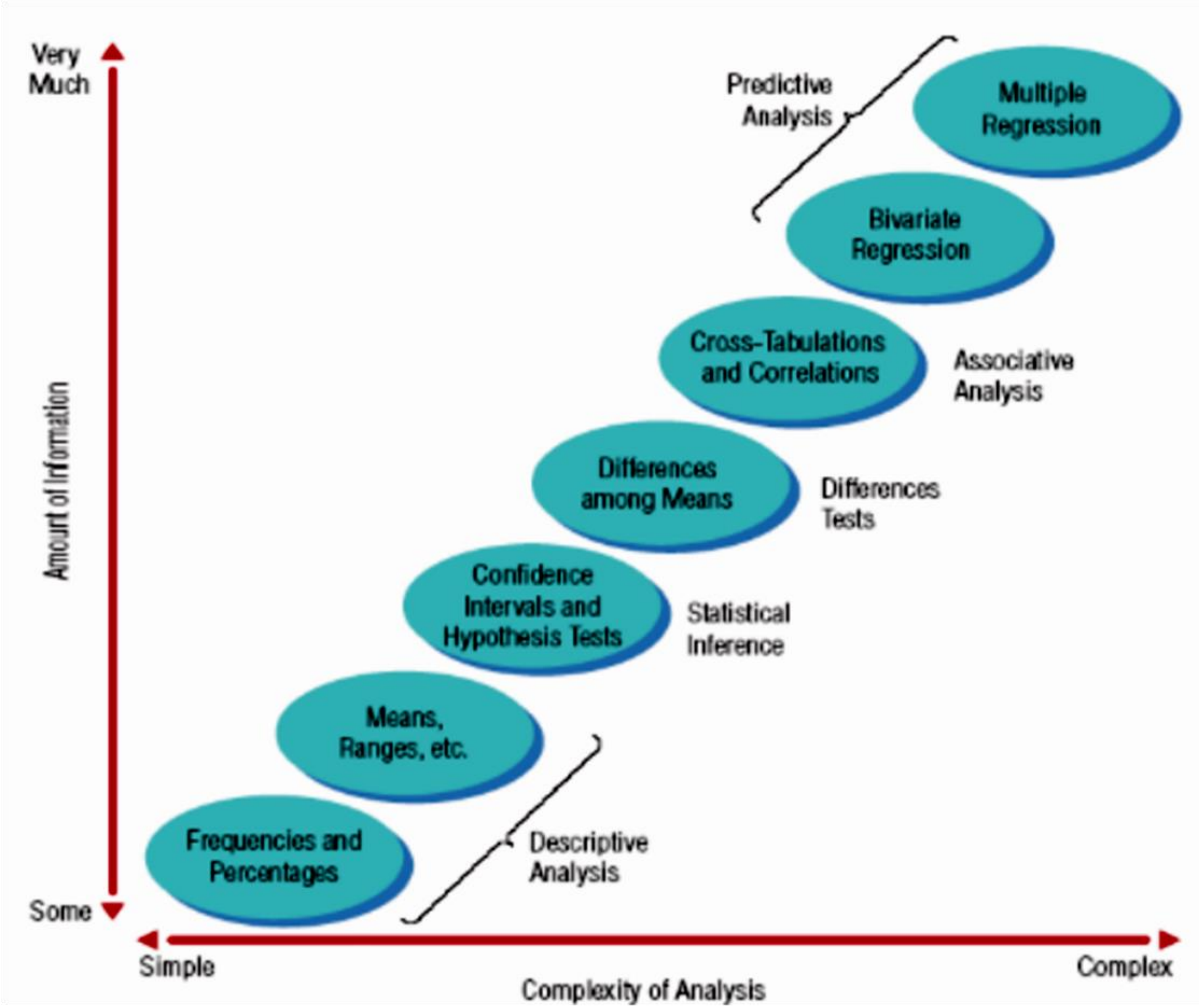
Saldana, J. (2009). The coding manual for qualitative researchers. Los Angeles, CA: Sage.

SPSS Lecture

The Research Cycle



Types of data analysis



What is SPSS

- ❖Originally SPSS Stands for Statistical Package for the Social Science
- ❖Now it stands for Statistical Product and Service Solutions
- ❖One of the most popular statistical packages which can perform highly complex data manipulation and analysis with simple instructions

Why use SPSS

- Describing data
  - Summarizing data values in graphical displays
    - Eg:- Mean, Median, Mode, Standard deviation etc.
- Testing Hypothesis
  - Drawing conclusions
    - Eg:- T-tests, Chi-square test, correlation etc.
- Examining Relationships
  - Predicting values with a mathematical model
    - Eg:-Linear regression etc.

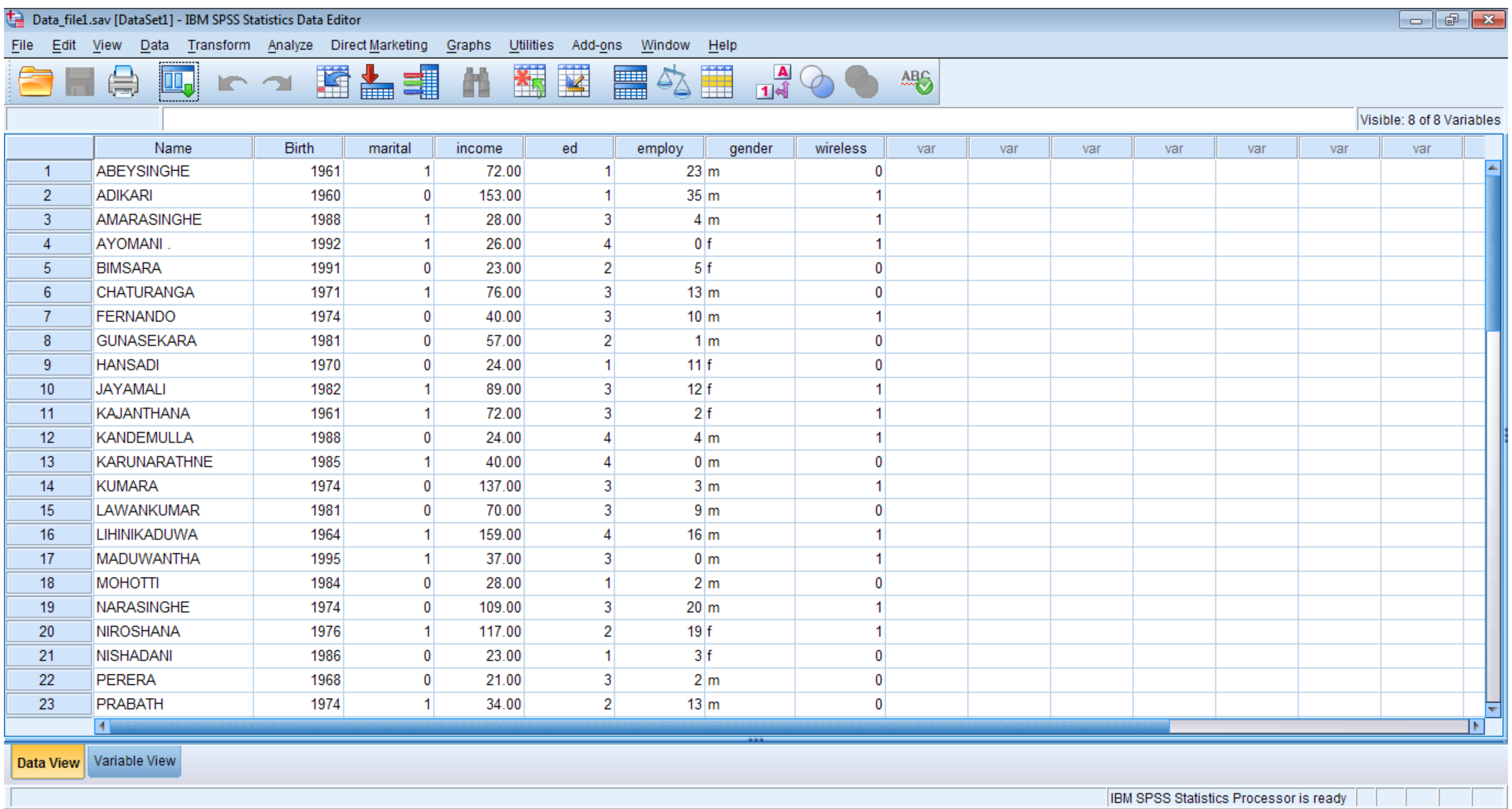
Windows of SPSS

- It has four windows
  - Data editor
  - Output viewer
  - Syntax editor
  - Script window

SPSS Windows: Data Editor window

- ❖ Data Editor

  Spreadsheet-like system for defining, entering, editing, and displaying data. Extension of the saved file will be “sav.”

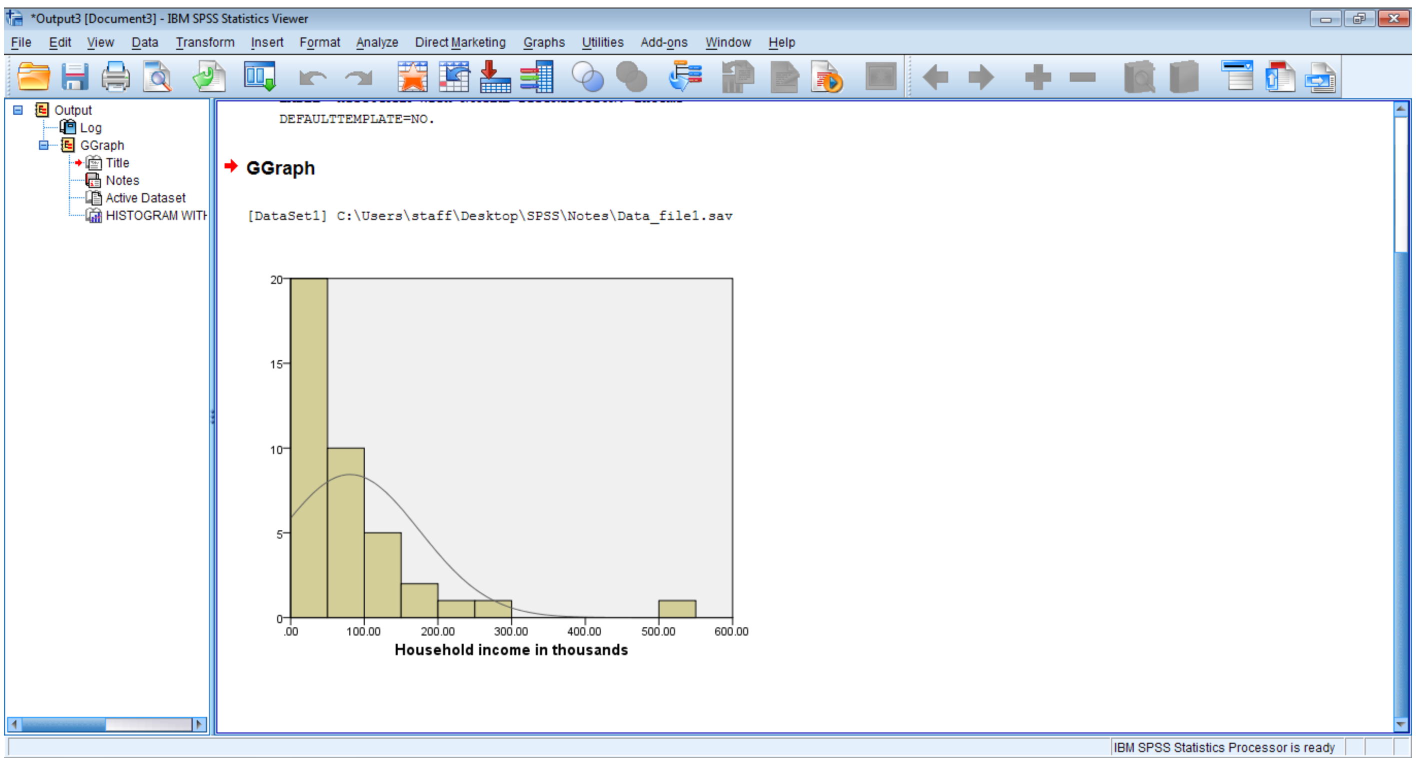


	Name	Birth	marital	income	ed	employ	gender	wireless	var	var	var	var	var	var	var
1	ABEYSINGHE	1961	1	72.00	1	23	m	0							
2	ADIKARI	1960	0	153.00	1	35	m	1							
3	AMARASINGHE	1988	1	28.00	3	4	m	1							
4	AYOMANI .	1992	1	26.00	4	0	f	1							
5	BIMSARA	1991	0	23.00	2	5	f	0							
6	CHATURANGA	1971	1	76.00	3	13	m	0							
7	FERNANDO	1974	0	40.00	3	10	m	1							
8	GUNASEKARA	1981	0	57.00	2	1	m	0							
9	HANSADI	1970	0	24.00	1	11	f	0							
10	JAYAMALI	1982	1	89.00	3	12	f	1							
11	KAJANTHANA	1961	1	72.00	3	2	f	1							
12	KANDEMULLA	1988	0	24.00	4	4	m	1							
13	KARUNARATHNE	1985	1	40.00	4	0	m	0							
14	KUMARA	1974	0	137.00	3	3	m	1							
15	LAWANKUMAR	1981	0	70.00	3	9	m	0							
16	LIHINIKADUWA	1964	1	159.00	4	16	m	1							
17	MADUWANTHA	1995	1	37.00	3	0	m	1							
18	MOHOTTI	1984	0	28.00	1	2	m	0							
19	NARASINGHE	1974	0	109.00	3	20	m	1							
20	NIROSHANA	1976	1	117.00	2	19	f	1							
21	NISHADANI	1986	0	23.00	1	3	f	0							
22	PERERA	1968	0	21.00	3	2	m	0							
23	PRABATH	1974	1	34.00	2	13	m	0							

SPSS Windows: Output Viewer

- Output Viewer

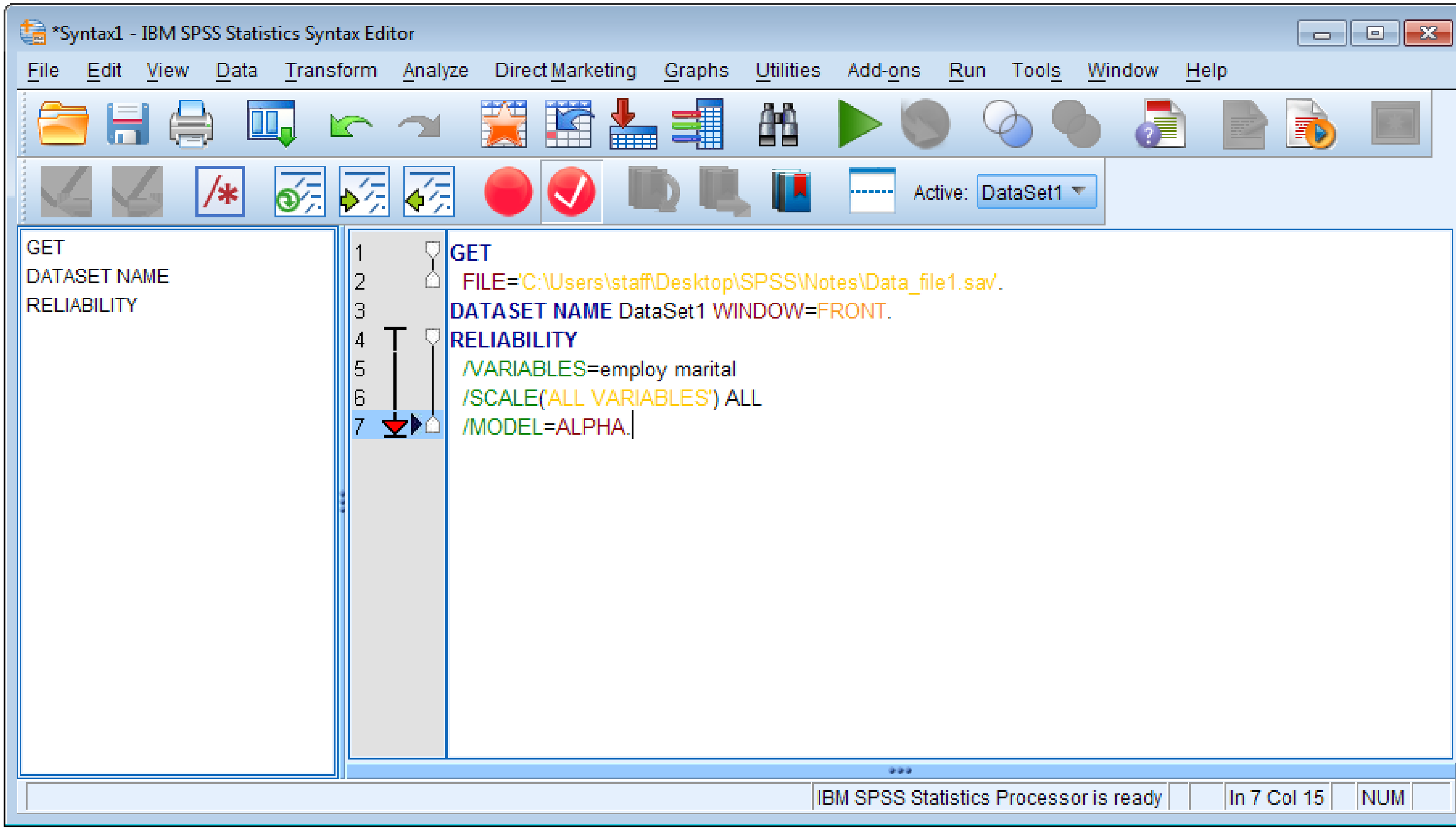
  Displays output and errors. Extension of the saved file will be “spv.”



SPSS Windows: Syntax Editor

- Syntax Editor

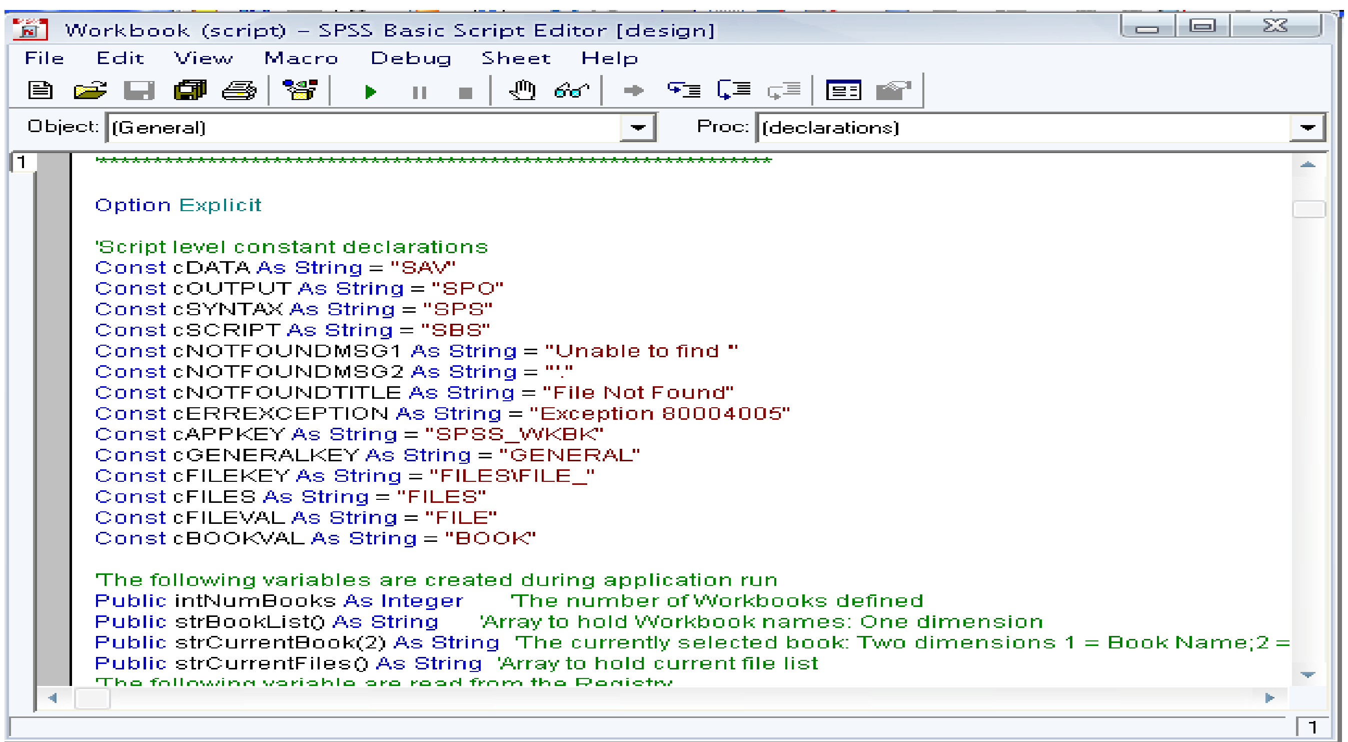
  Text editor for syntax composition. Extension of the saved file will be “sps.”



SPSS Windows: Script Window

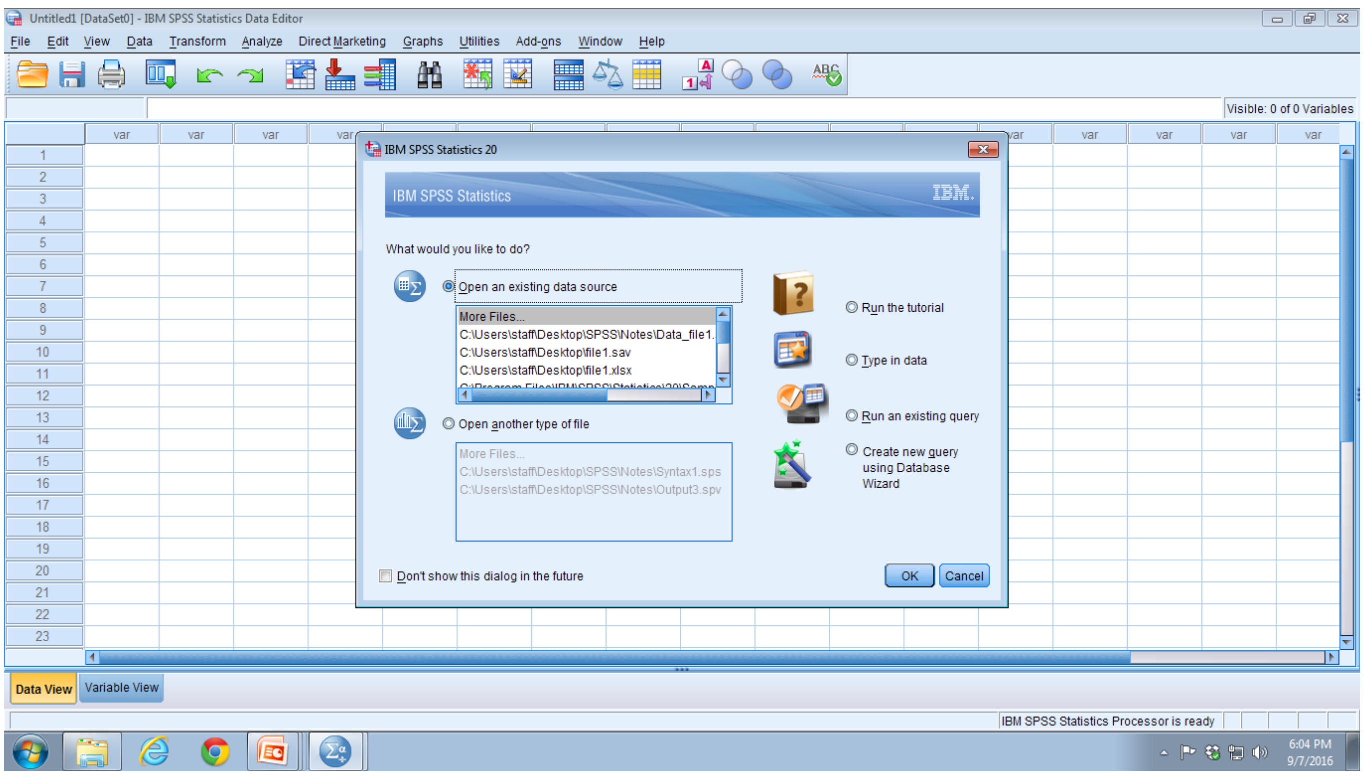
- Script Window

  Provides the opportunity to write full-blown programs, in a BASIC-like language. Text editor for syntax composition. Extension of the saved file will be “sbs.”



Opening SPSS

- Open SPSS can be done in one of the following ways
  - Click on the desktop shortcut
  - Start → All Programmes → IBM SPSS Statistics → IBM SPSS Statistics 20
  - Click on IBM SPSS Statistics 20 in start menu or frequently used programme list after clicking on the start button



Entering Data

- Manually Enter Data:

  1. Define Variables in Variable View
  2. Enter data in Data view

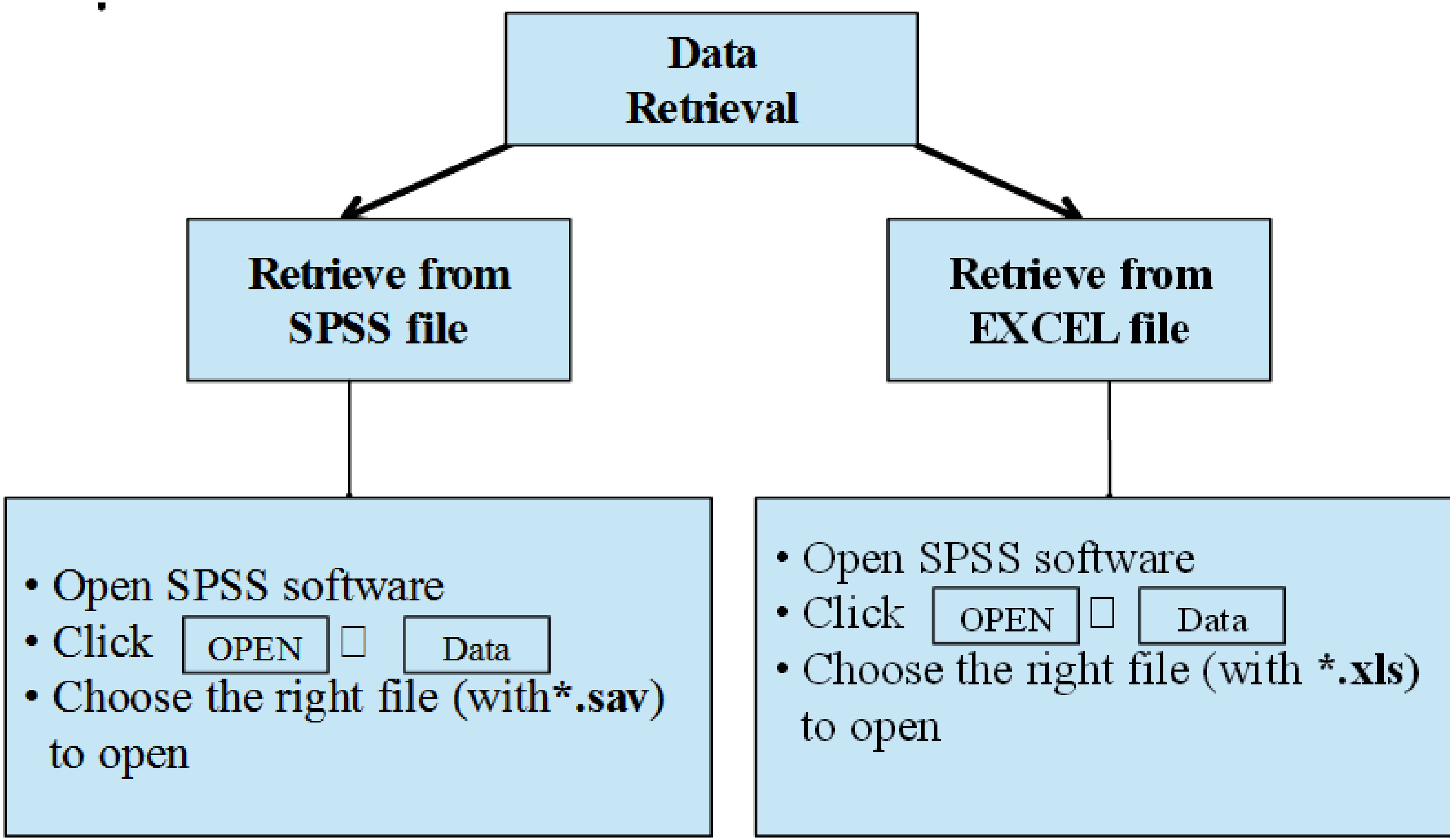
  Or

- Using the menu select **“File”** and then **“Open”** an existing SPSS data document (.sav)

  Or

- Read Data in from Excel

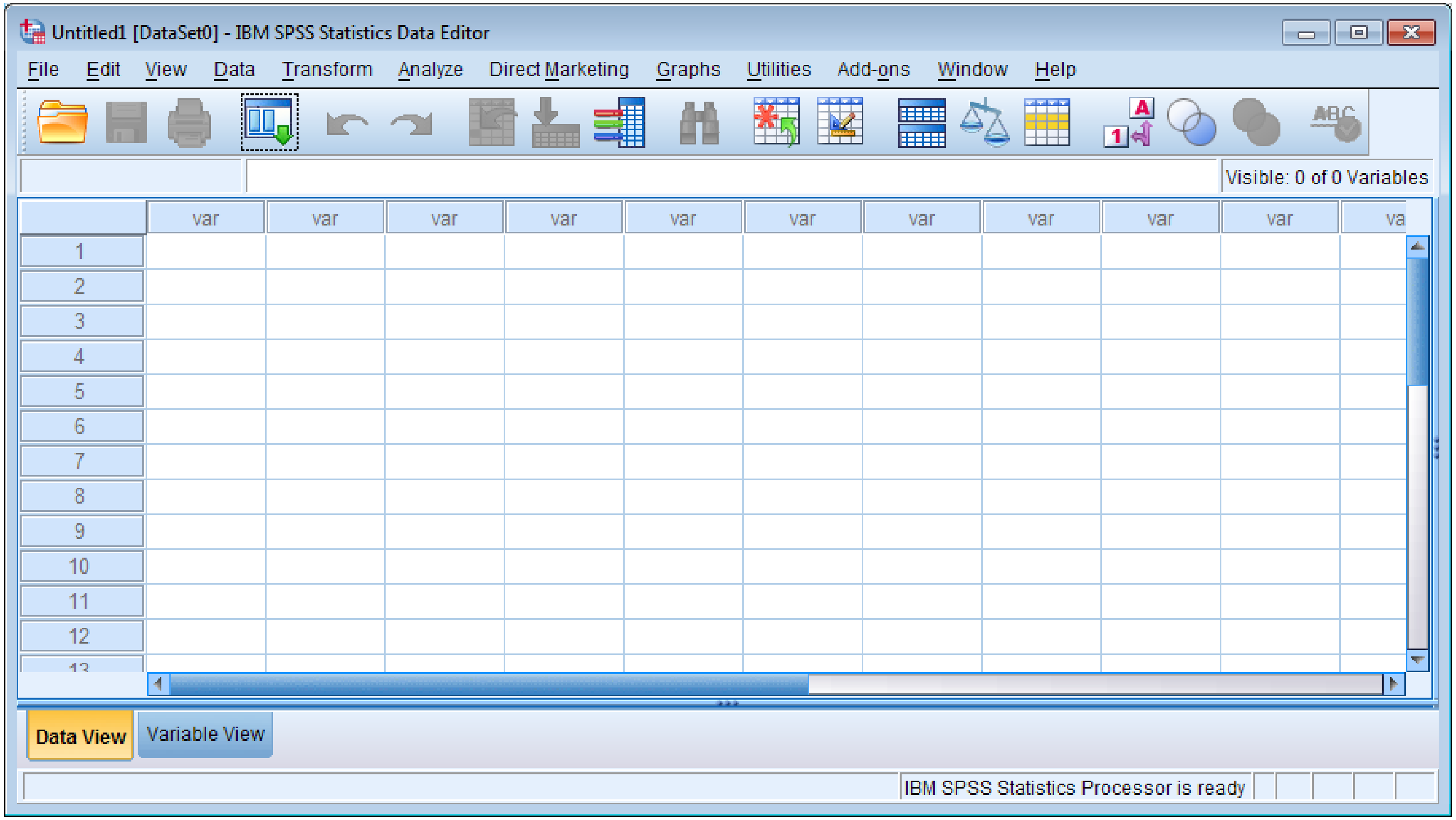
Data retrieve



SPSS Window

- The default window will have the data editor
- There are two sheets in the window:

  1. Data view   2. Variable view

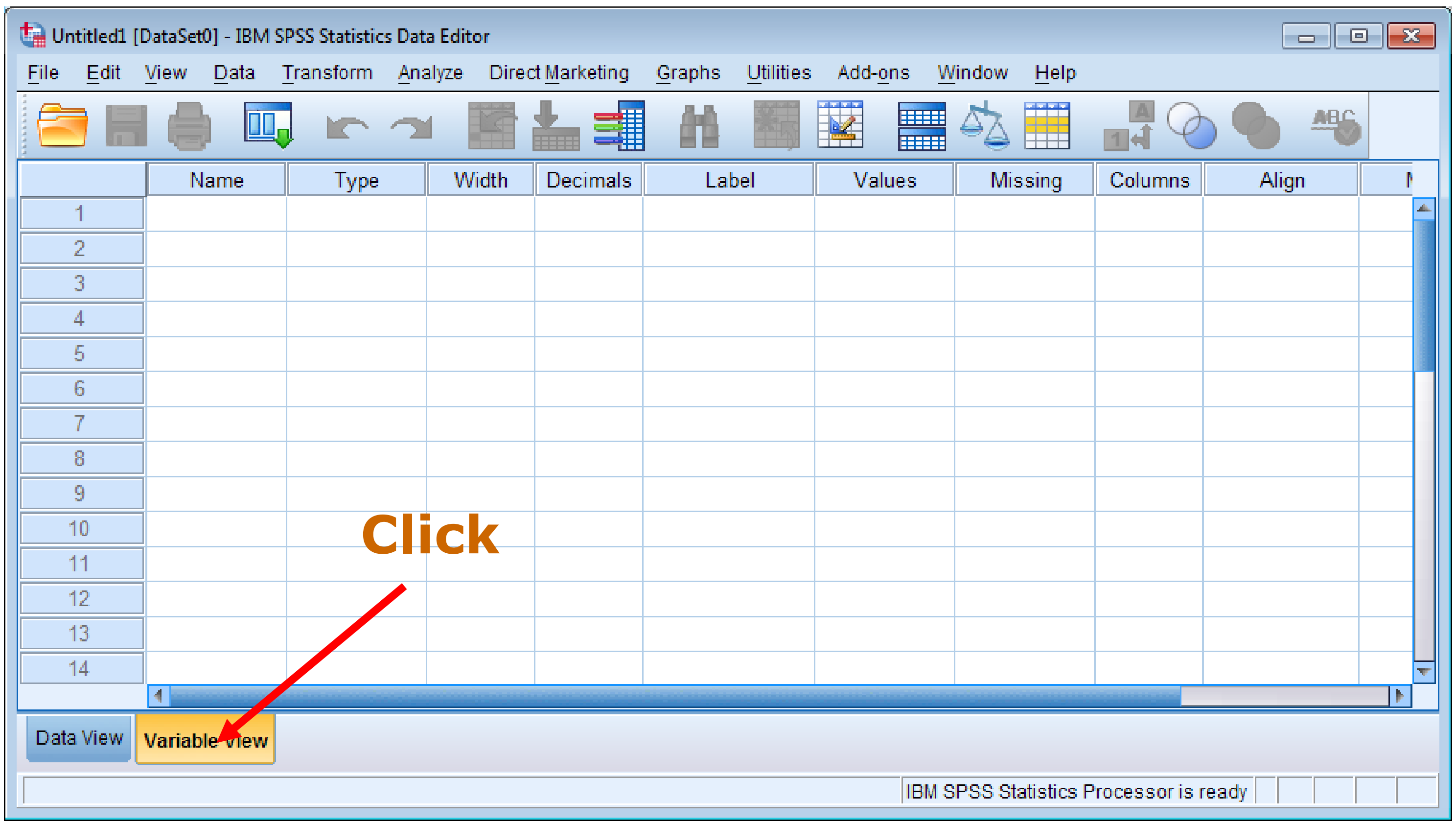


Data View Window

- The Data View window

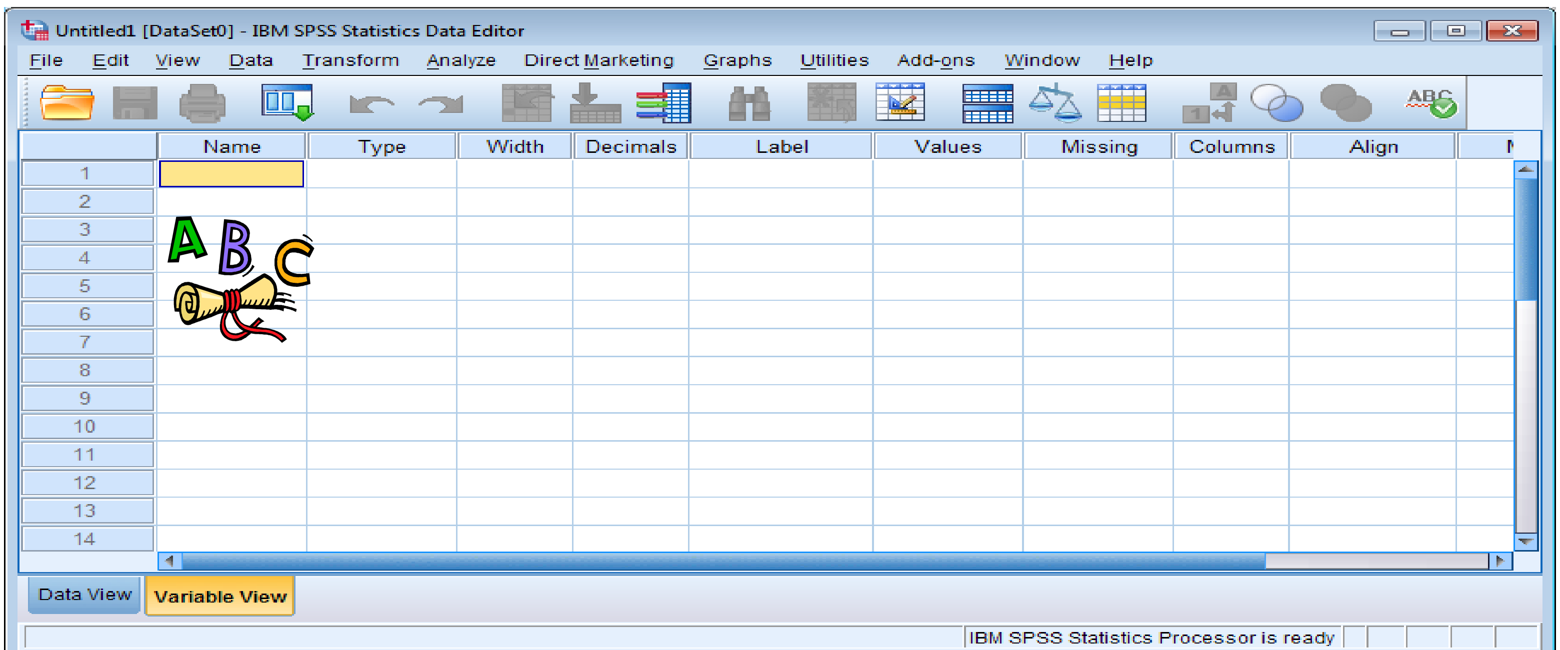
  This sheet is visible when you first open the Data Editor and this sheet contains the data

- Click on the tab labeled Variable View



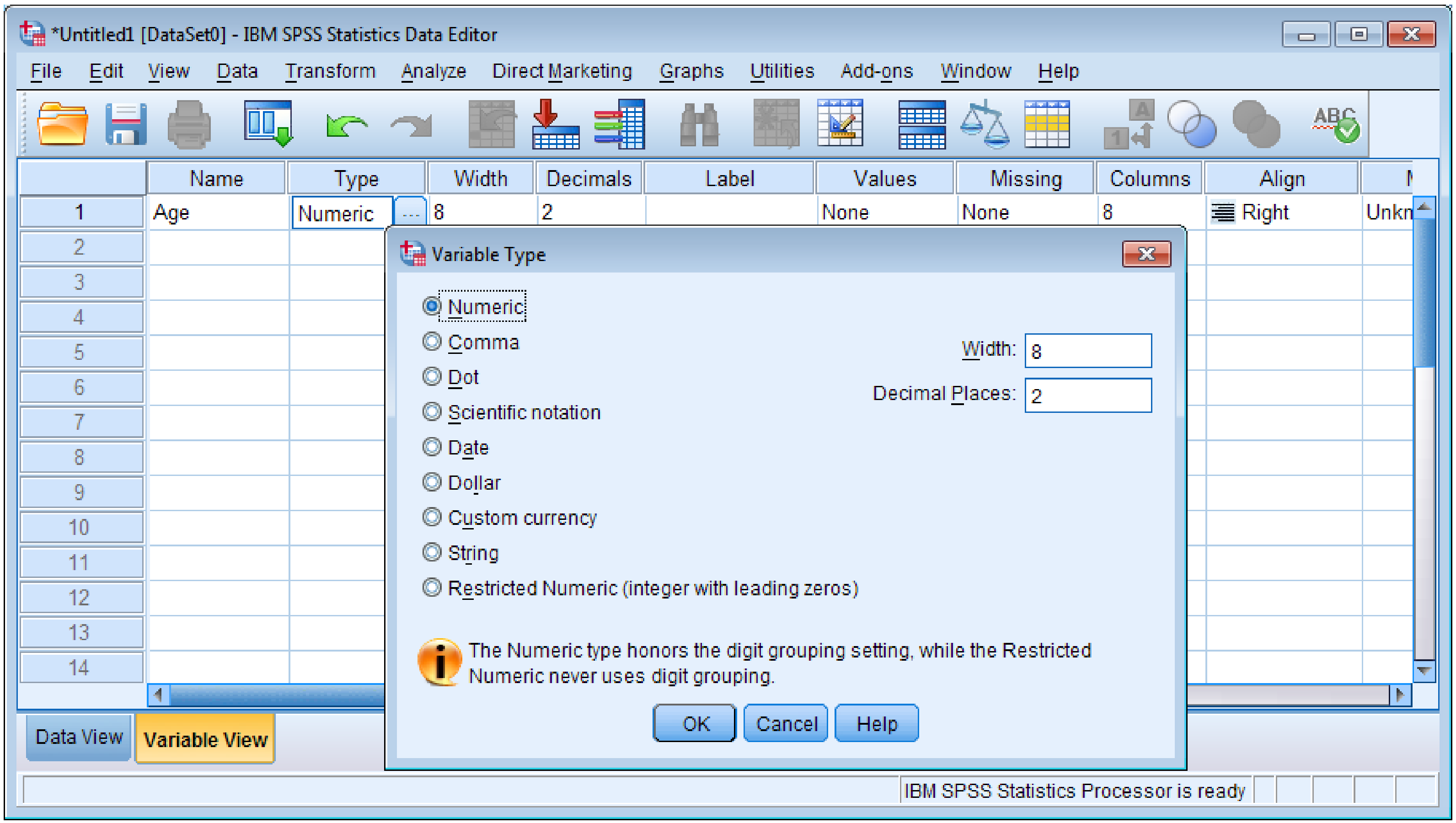
Variable View Window

- This sheet contains information about the data set that is stored with the dataset
- Name
  - The first character of the variable name must be alphabetic
  - Variable names must be unique, and have to be less than 64 characters.
  - Spaces are NOT allowed.



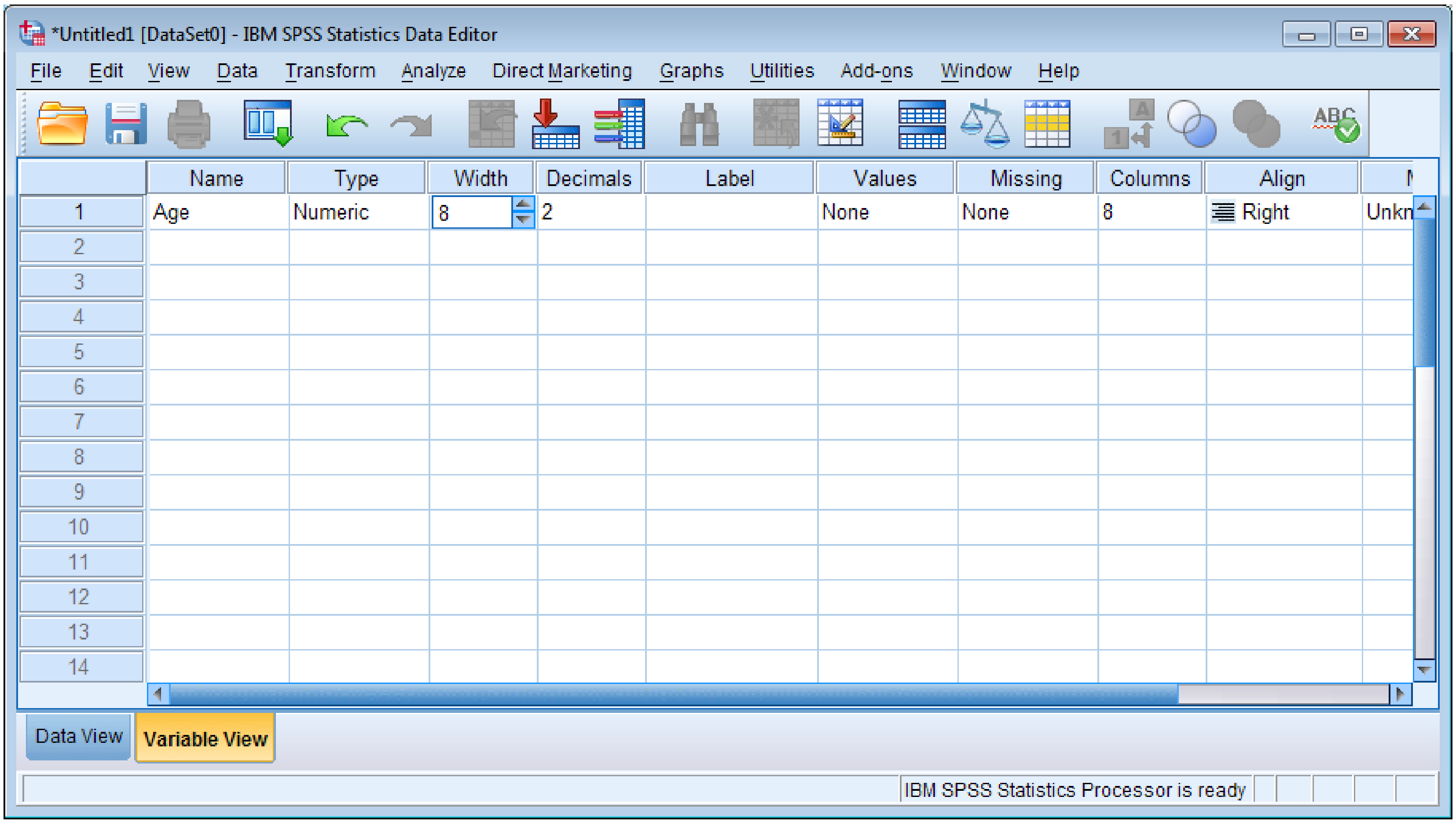
Variable View Window: Type

- Type
  - Click on the ‘type’ box. The two basic types of variables that you will use are numeric and string. This column enables you to specify the type of variable.



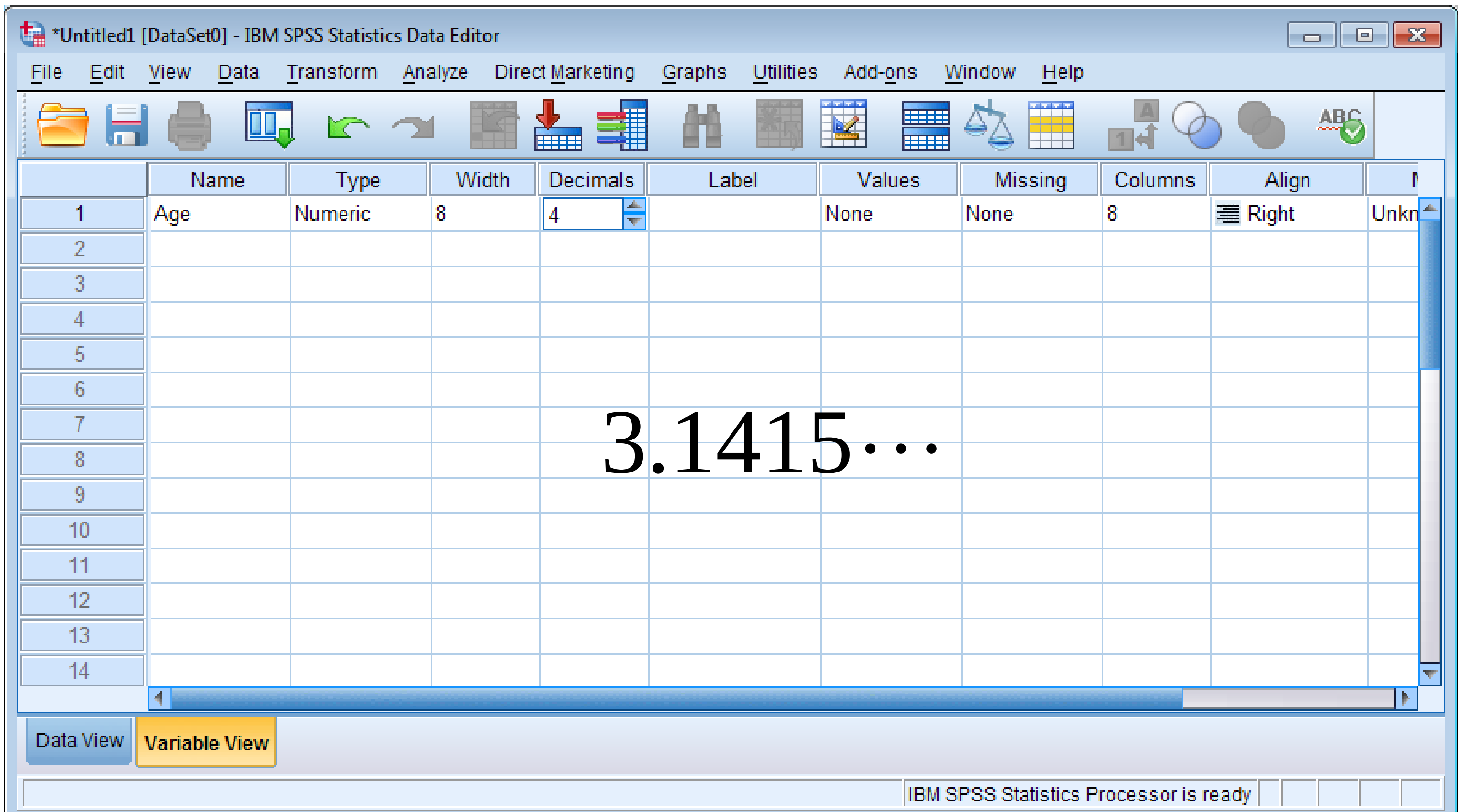
Variable View Window: Width

- Width
  - Width allows you to determine the number of characters SPSS will allow to be entered for the variable



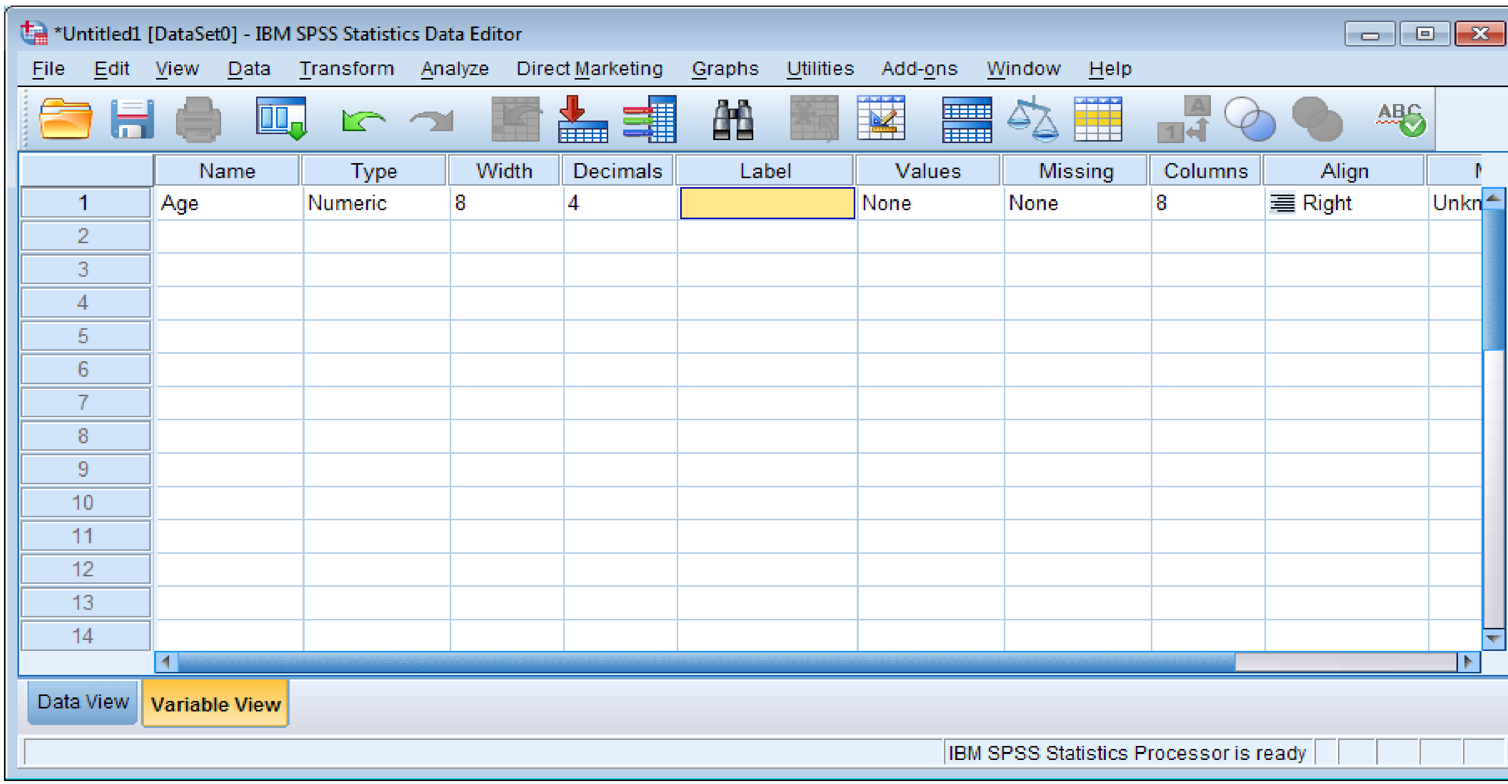
Variable View Window: Decimals

- Decimals
  - Number of decimals
  - It has to be less than or equal to 16



Variable View Window: Label

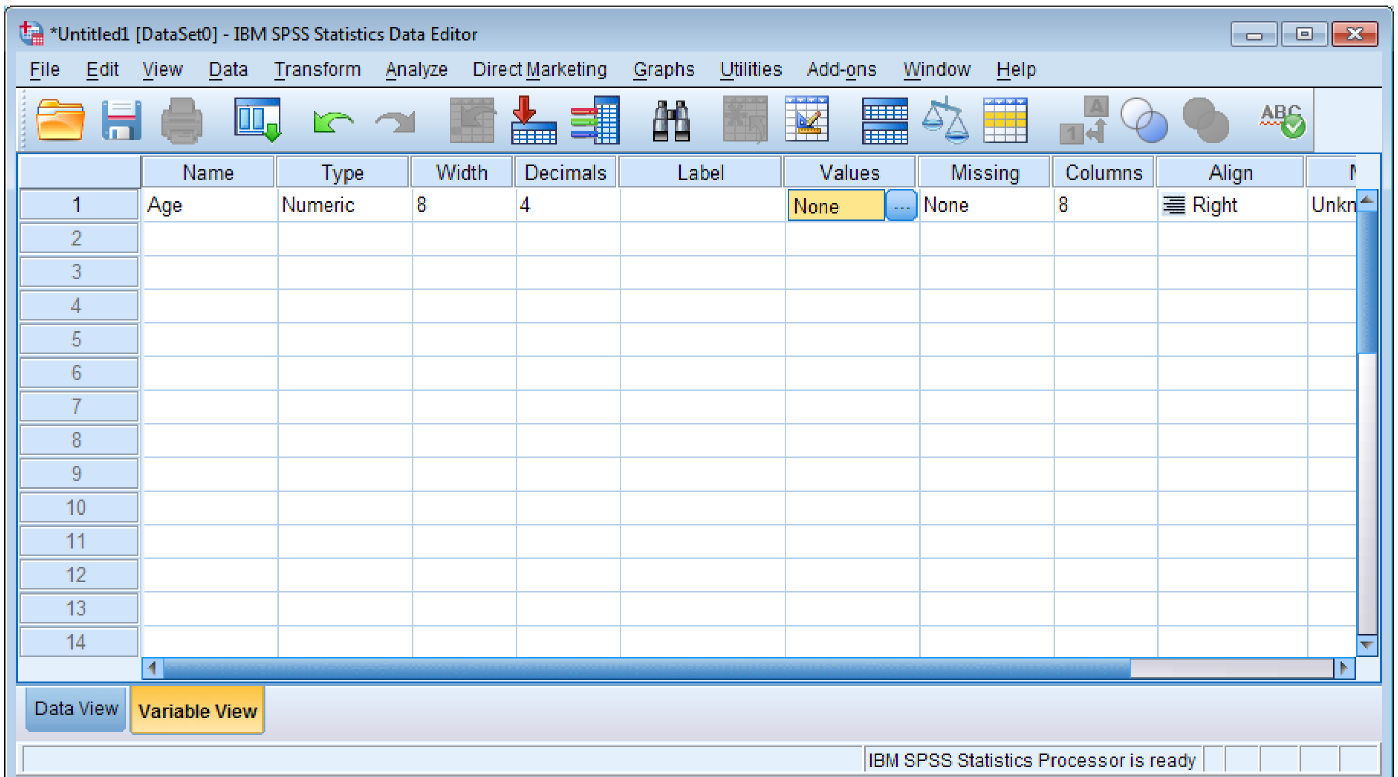
- Label
  - You can specify the details of the variable
  - You can write characters with spaces up to 256 characters



Variable View Window: Values

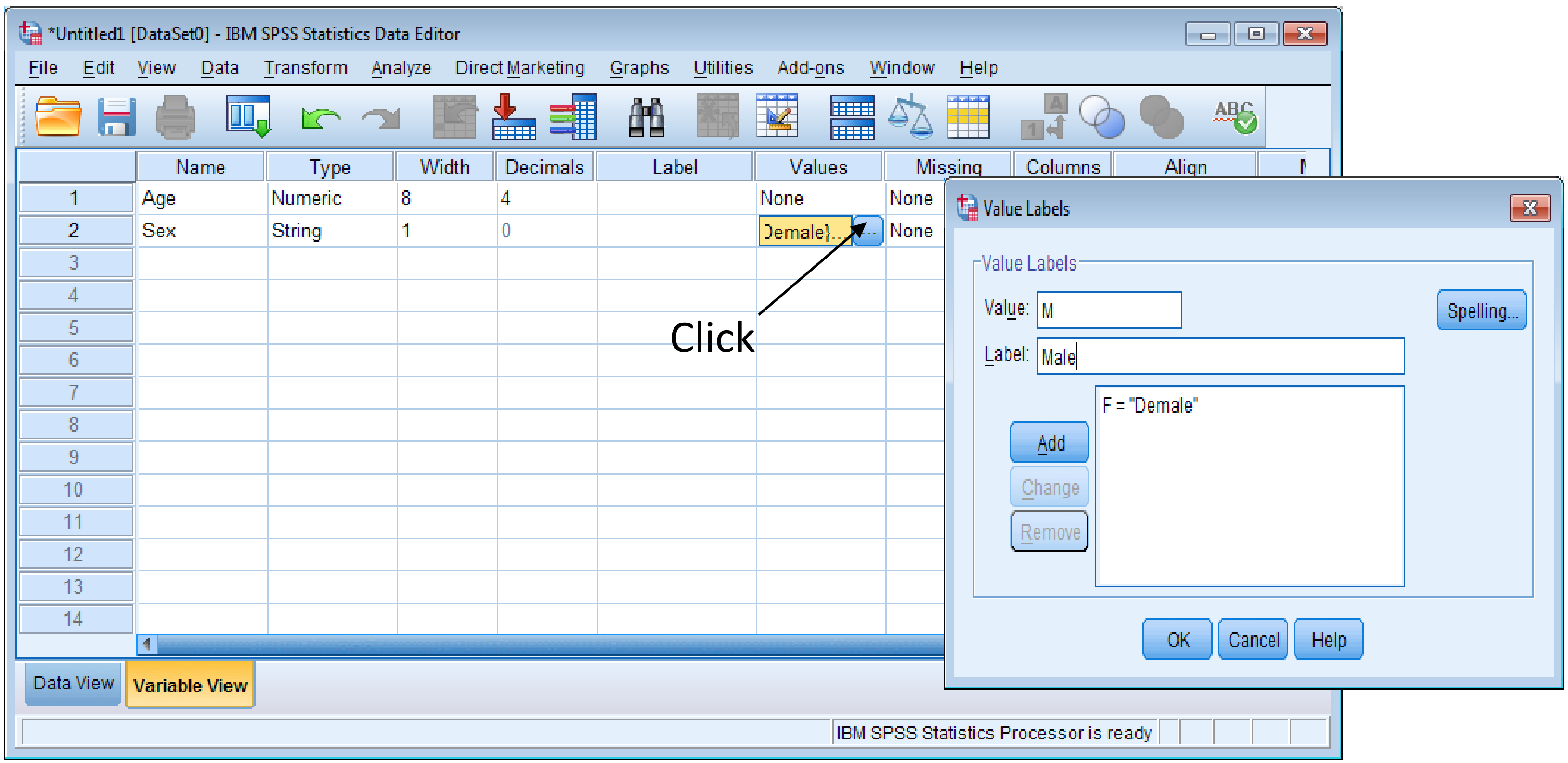
❖ Values

- ❑ This is used and to suggest which numbers represent which categories when the variable represents a category



Defining the value labels

- Click the cell in the values column as shown below
- For the value, and the label, you can put up to 60 characters.
- After defining the values click add and then click OK.



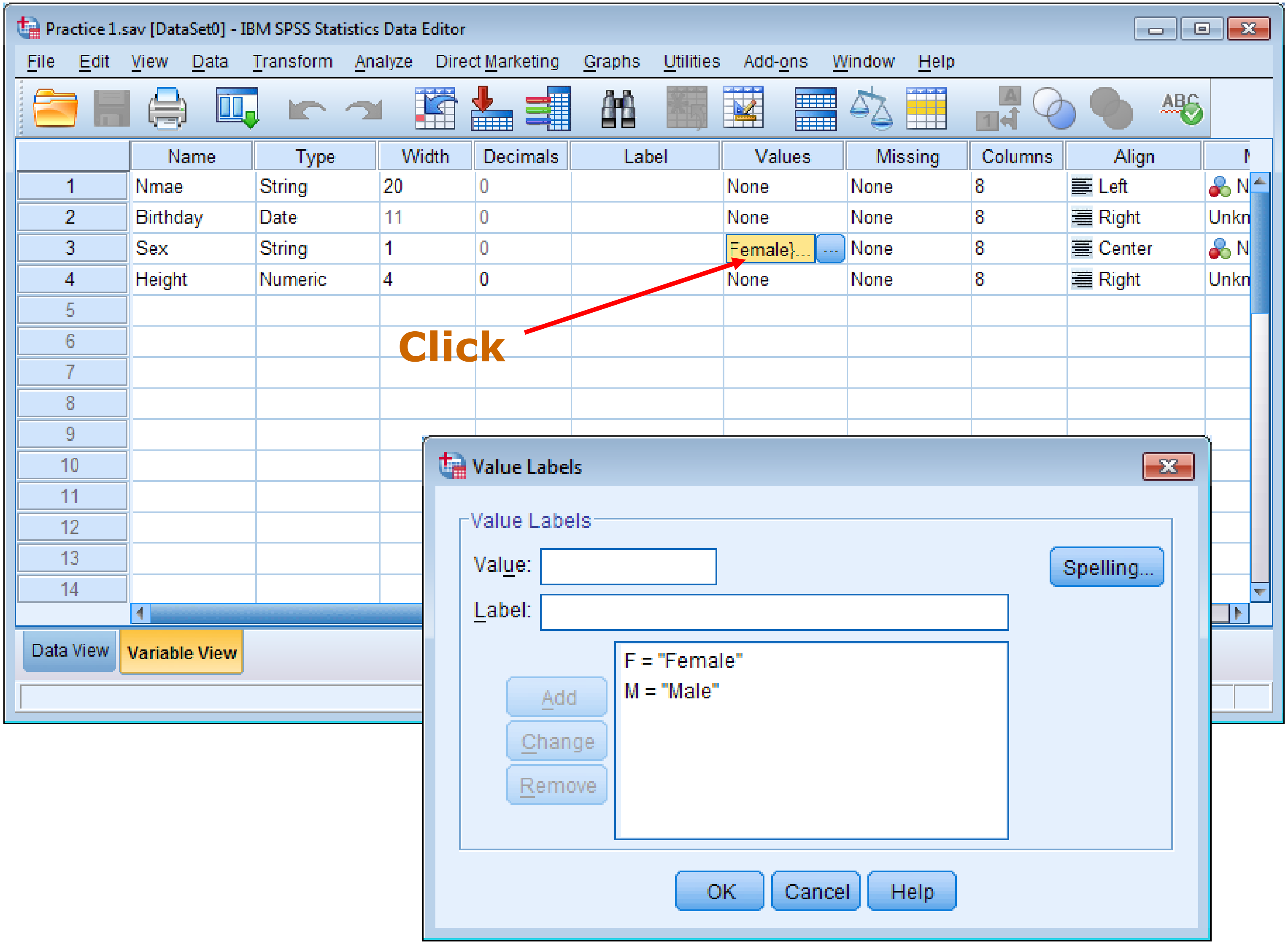
Practice 1

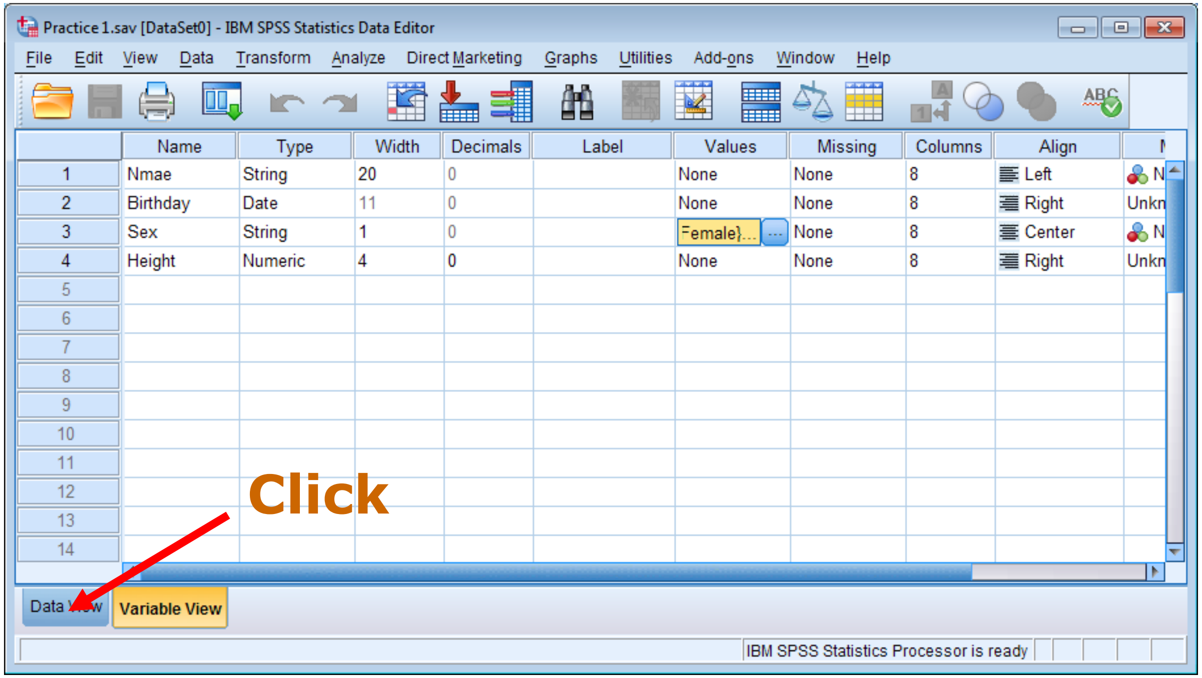
- Key the following information into SPSS?

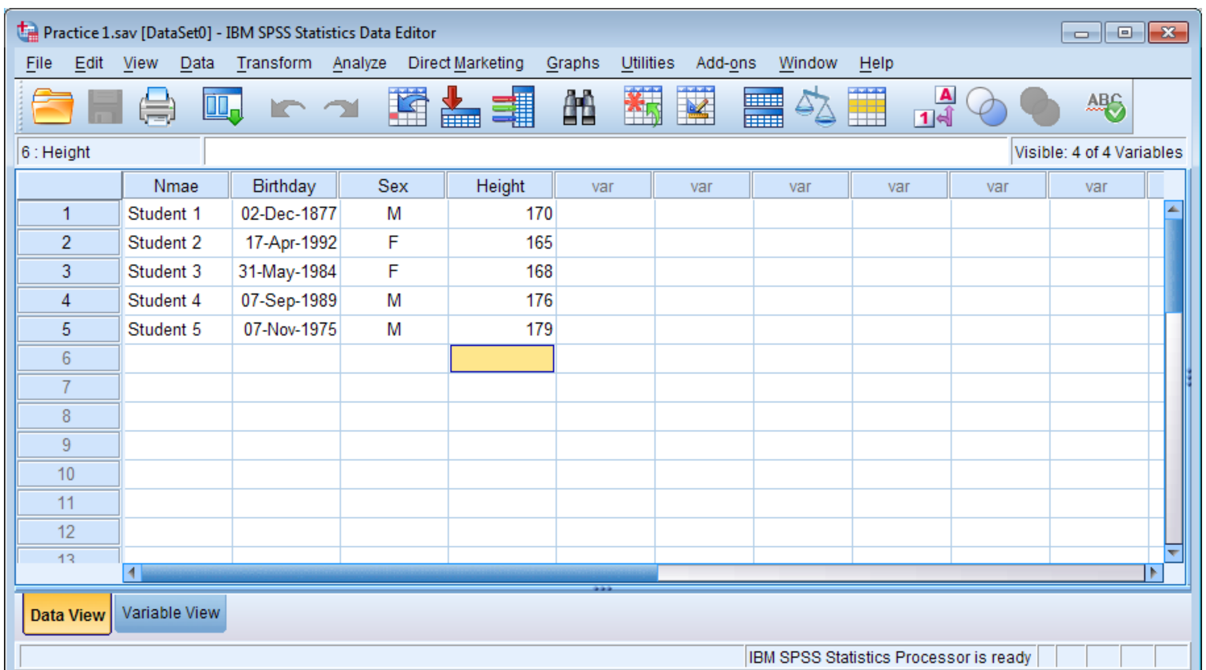
Name	Birthday	Sex	Height (cm)
Student 1	1977-12-02	M	170
Student 2	1992-04-17	F	165
Student 3	1984-05-31	F	168
Student 4	1989-09-07	M	176
Student 5	1975-11-07	M	179

Value = M represents Male and Value = F represents Female

Practice 1 (Solution Sample)

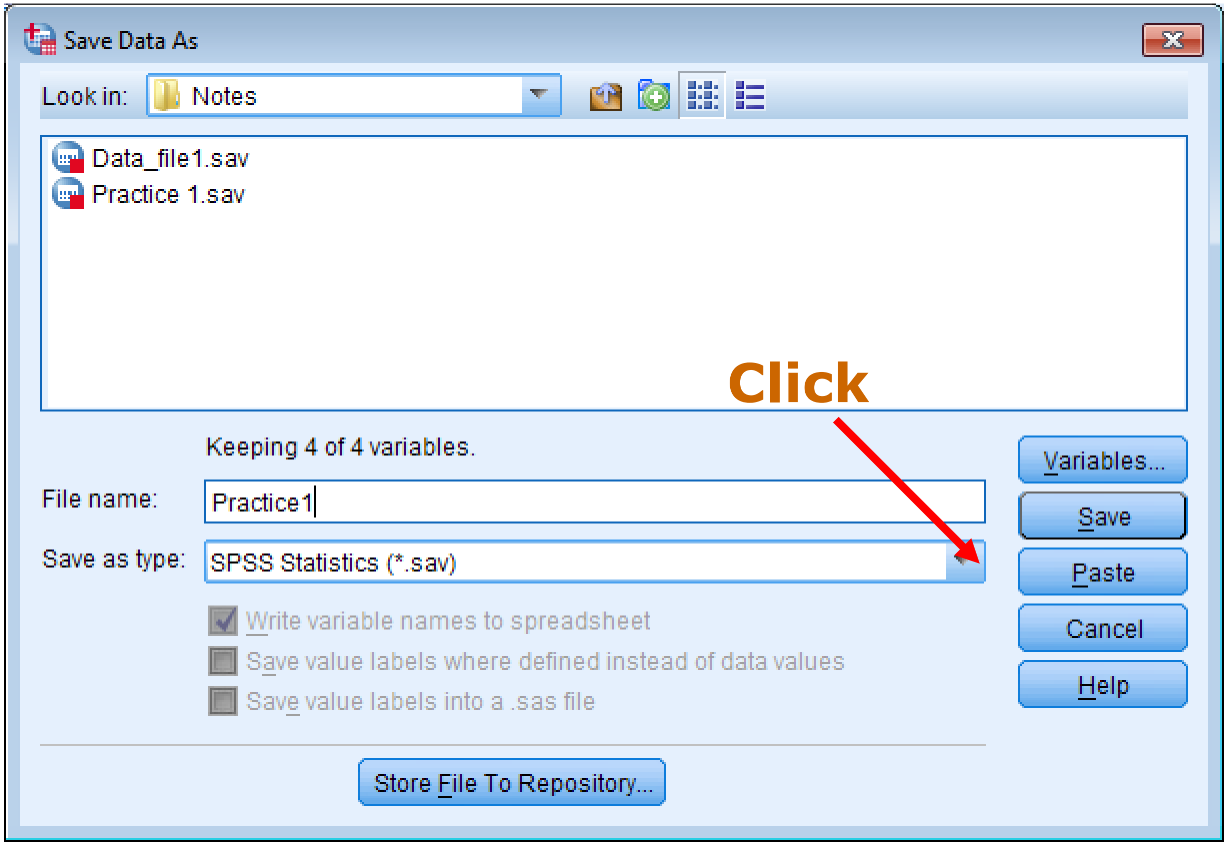



Practice 1.sav [DataSet0] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align
1 Nmae String 20 0 None None 8 Left
2 Birthday Date 11 0 None None 8 Right Unkn
3 Sex String 1 0 Female}... None 8 Center
4 Height Numeric 4 0 None None 8 Right Unkn
Click
Data View
Variable View
IBM SPSS Statistics Processor is ready


Practice 1.sav [DataSet0] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
6 : Height
Visible: 4 of 4 Variables
Nmae Birthday Sex Height var var var var var var
1 Student 1 02-Dec-1877 M 170
2 Student 2 17-Apr-1992 F 165
3 Student 3 31-May-1984 F 168
4 Student 4 07-Sep-1989 M 176
5 Student 5 07-Nov-1975 M 179
Data View
Variable View
IBM SPSS Statistics Processor is ready

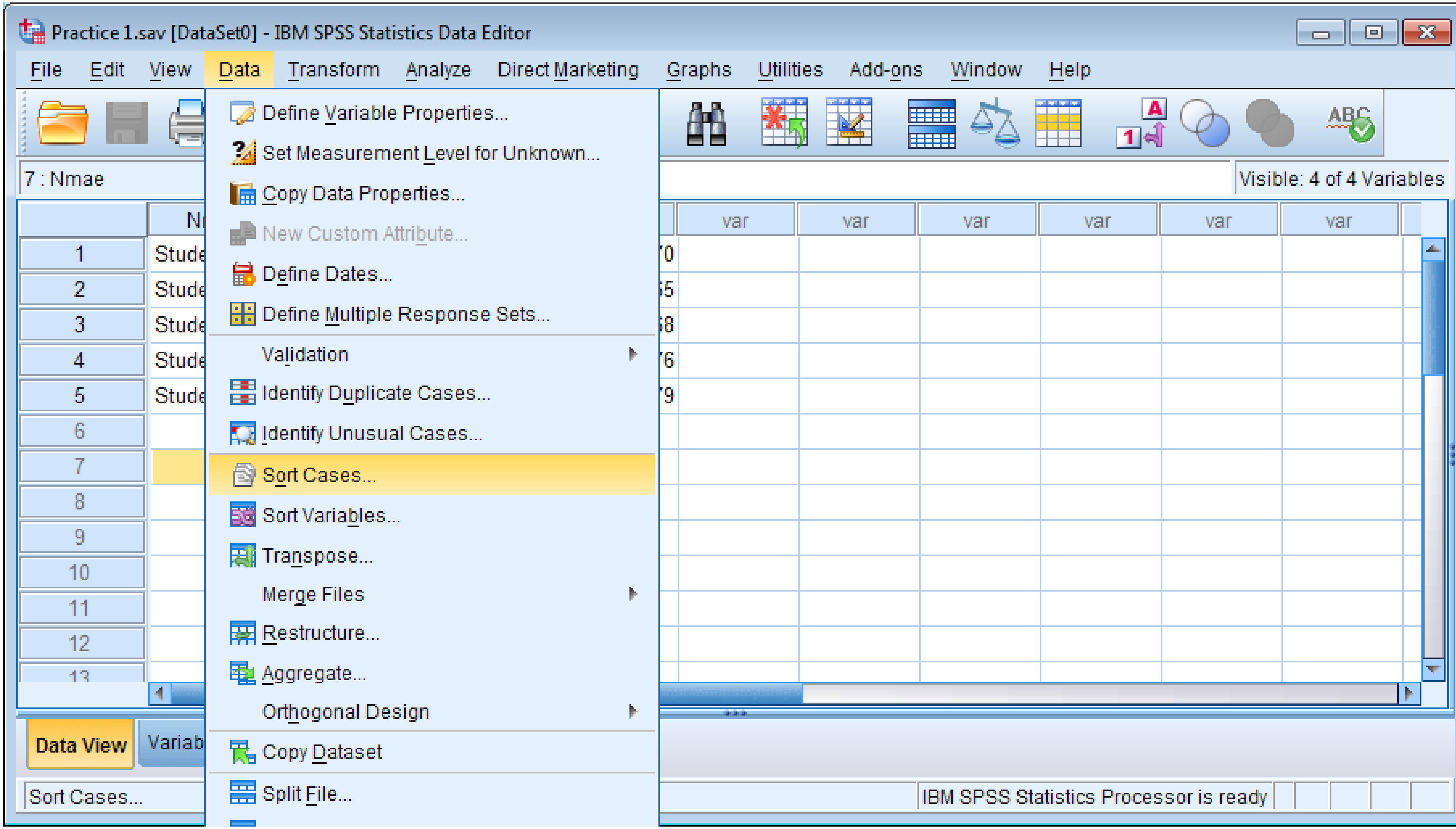
Saving the data

- To save the data file you created simply click 'file' and click 'save as.' You can save the file in different forms by clicking "Save as type."



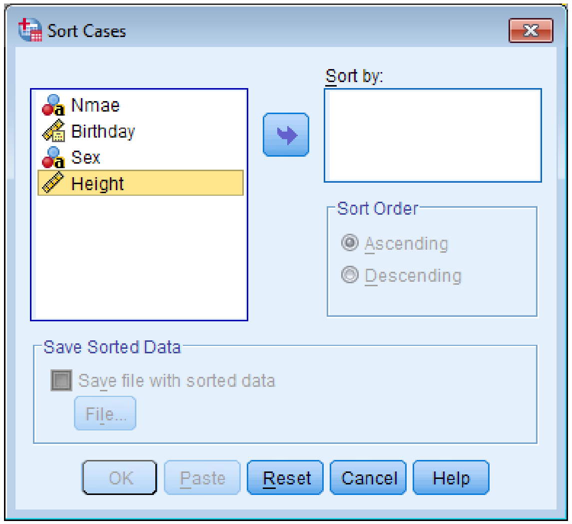
Sorting the data

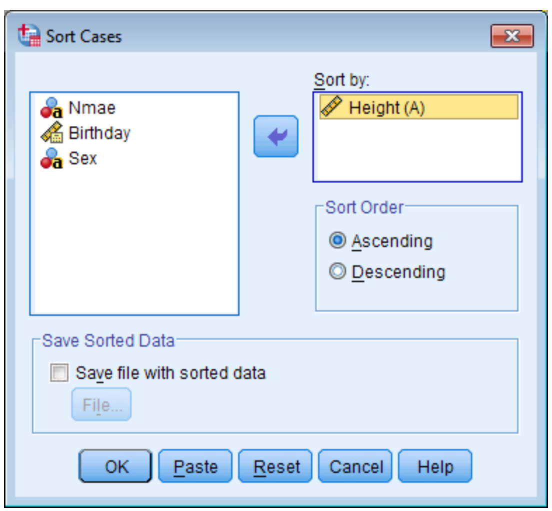
❖ Click ‘Data’ and then click Sort Cases

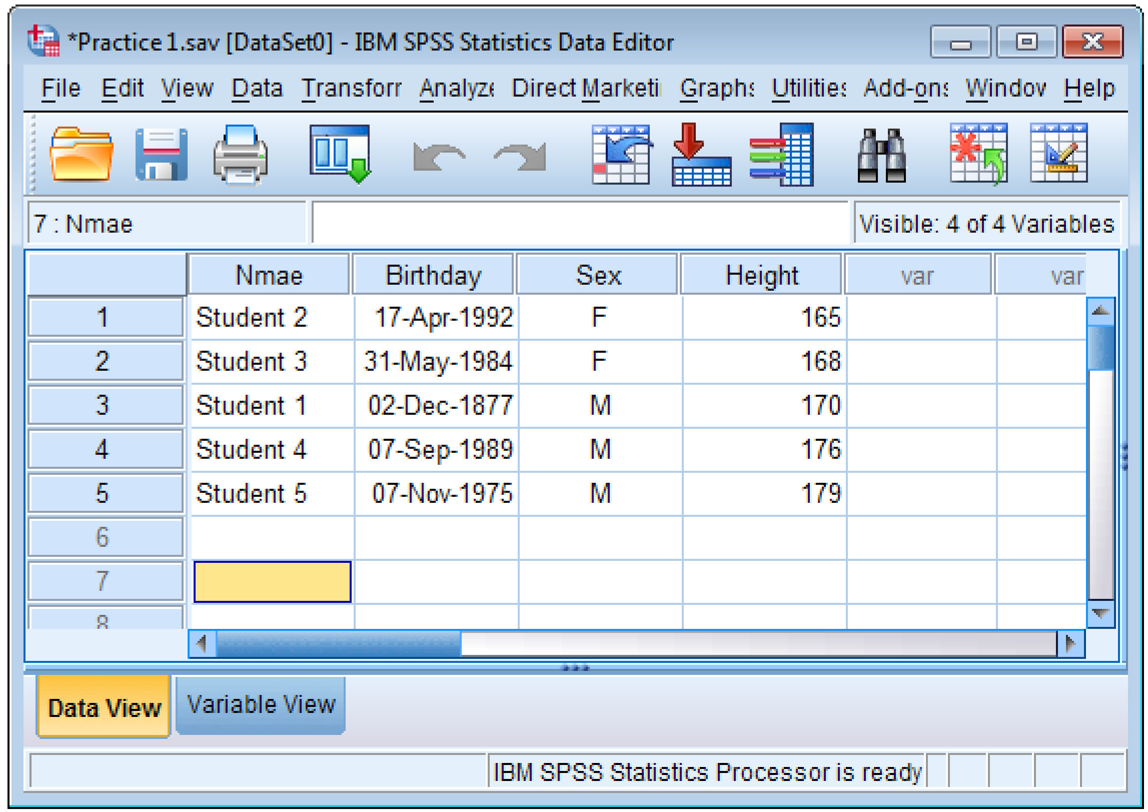


Sorting the data (cont'd)

❖ Double Click 'Height.' Then click ok.

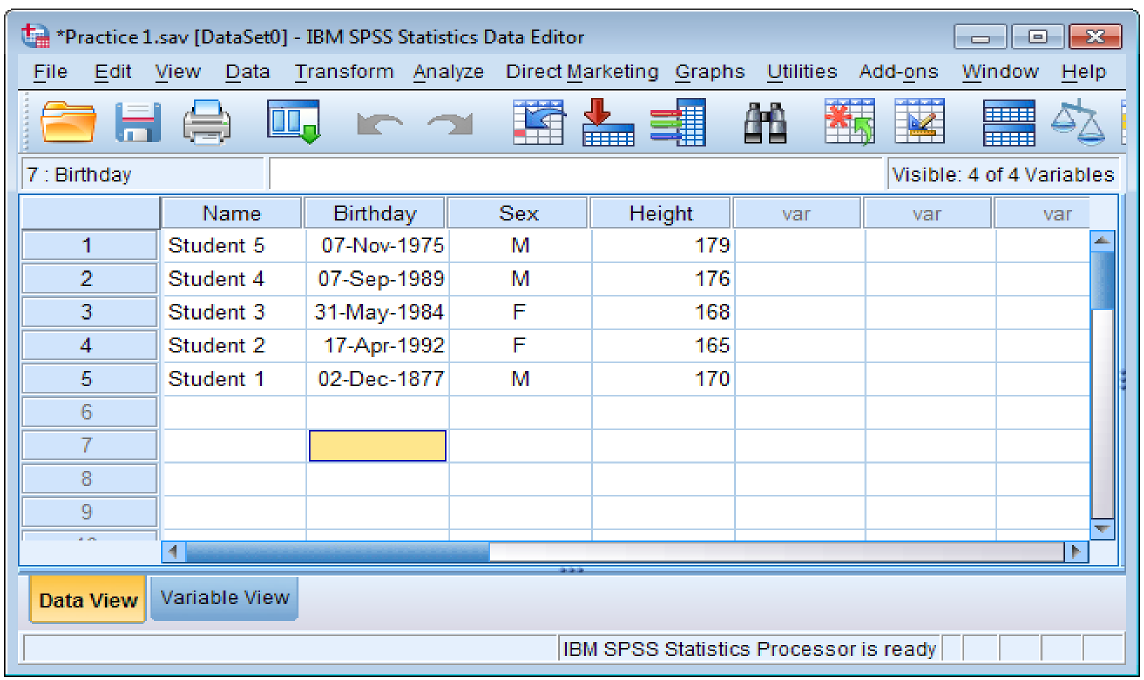






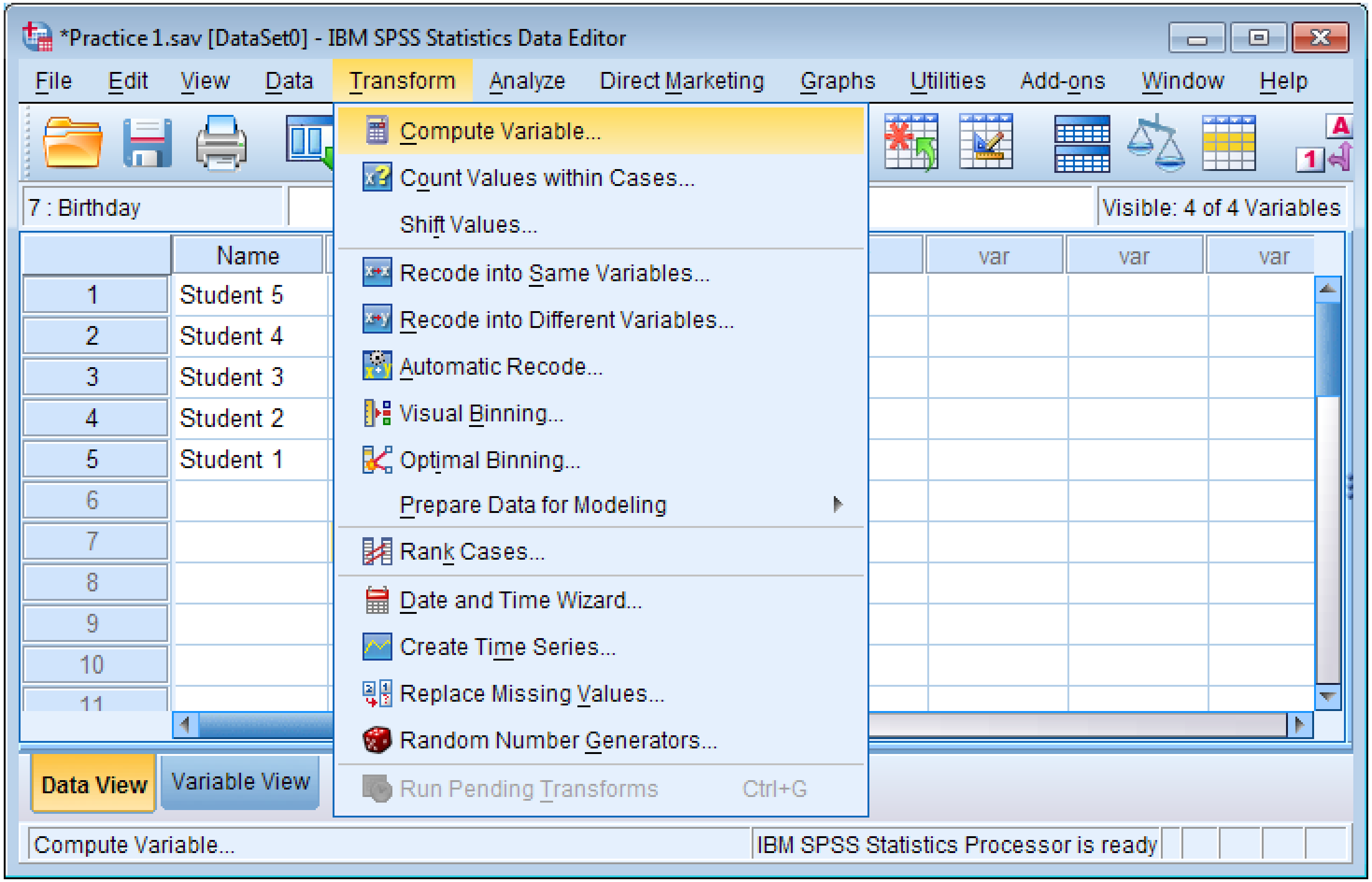
Practice 2

- How would you sort the data by the ‘Name’ of students in descending order?
- Answer
  - Click data, sort cases, double click ‘name of students,’ click ‘descending,’ and finally click ok.



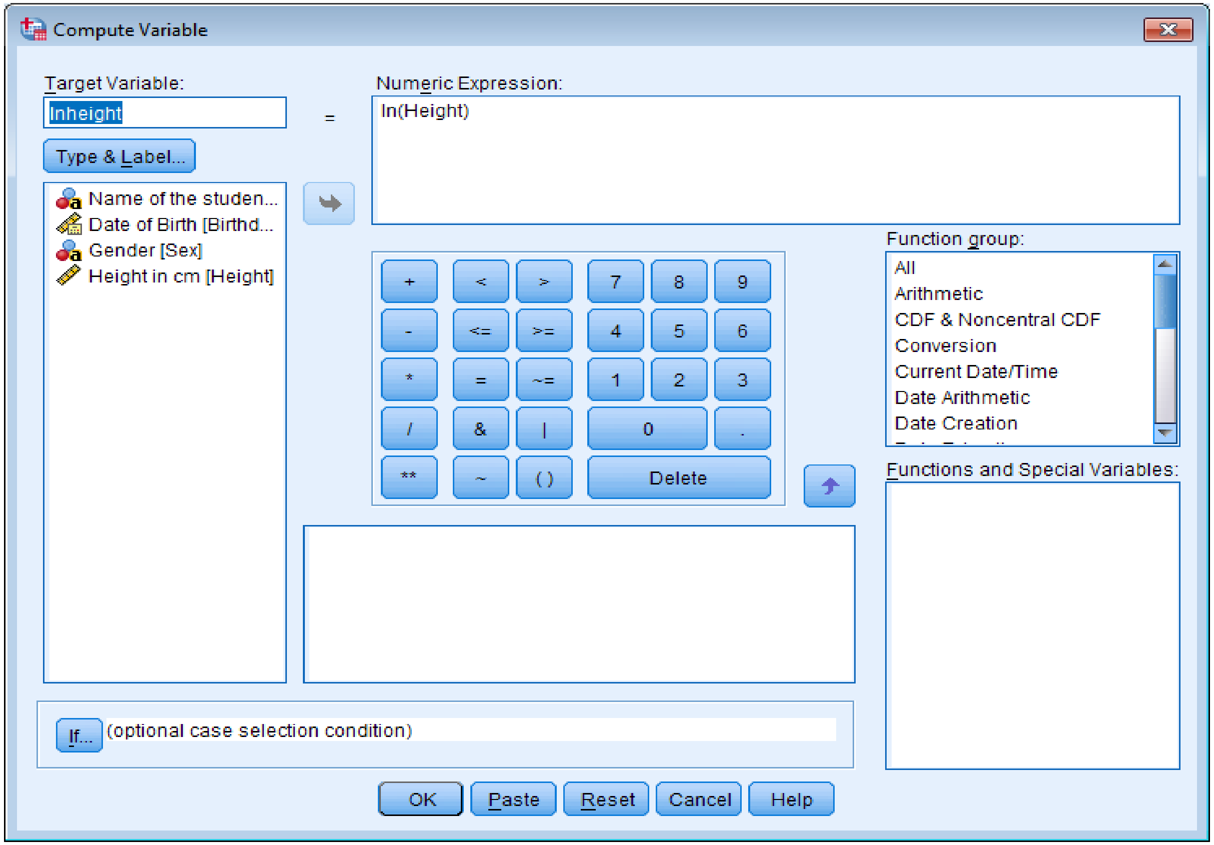
Transforming data

❖ Click ‘Transform’ and then click ‘Compute Variable…’



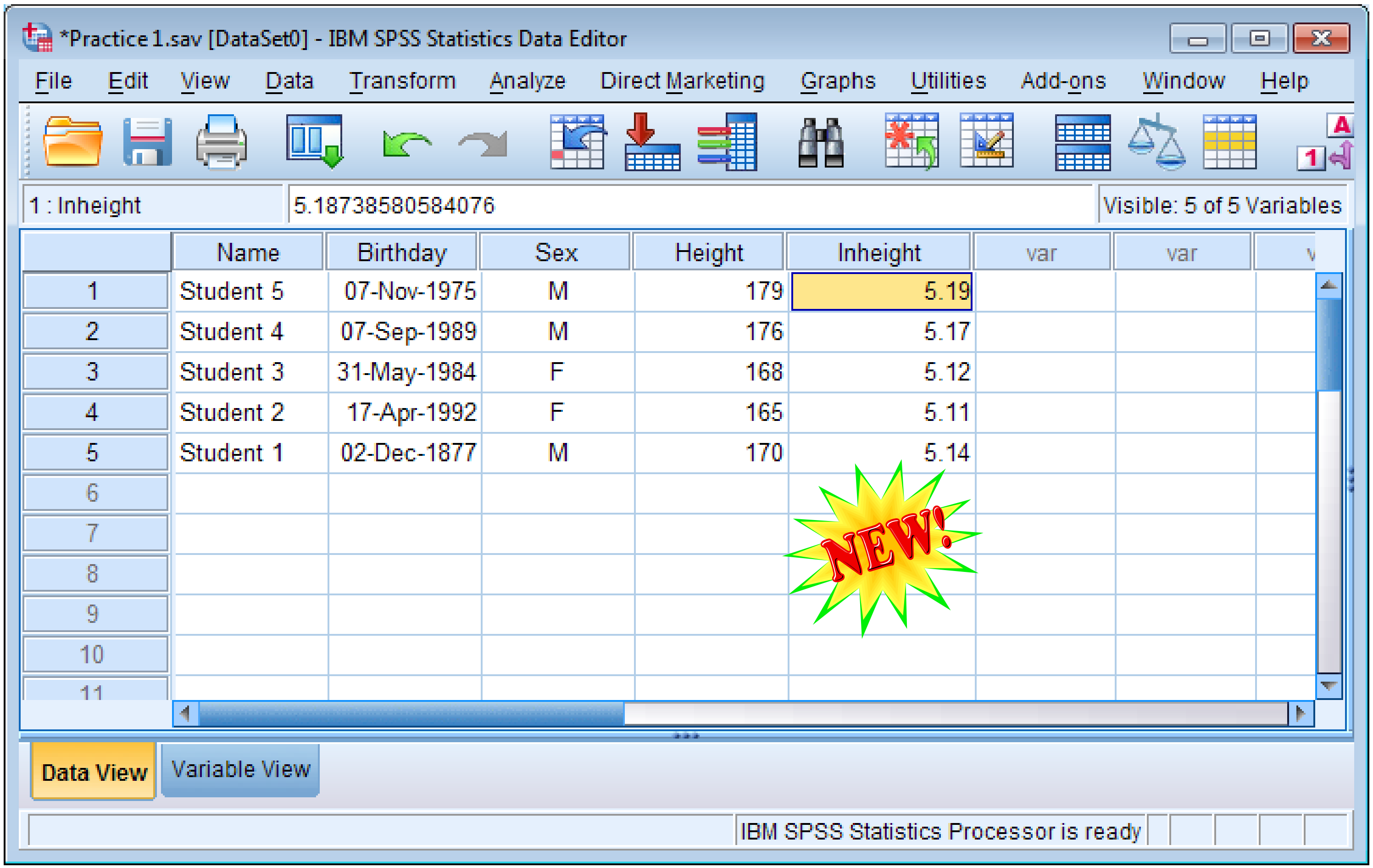
Transforming data (cont'd)

- Example: Adding a new variable named 'lnheight' which is the natural log of height
  - Type in lnheight in the 'Target Variable' box. Then type in 'ln(Height)' in the 'Numeric Expression' box. Click OK



Transforming data (cont'd)

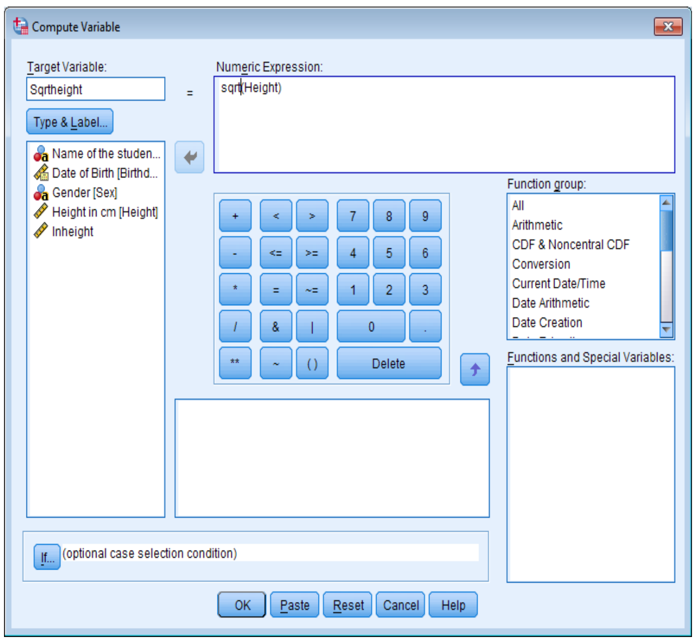
❖ A new variable 'lnheight' is added to the table

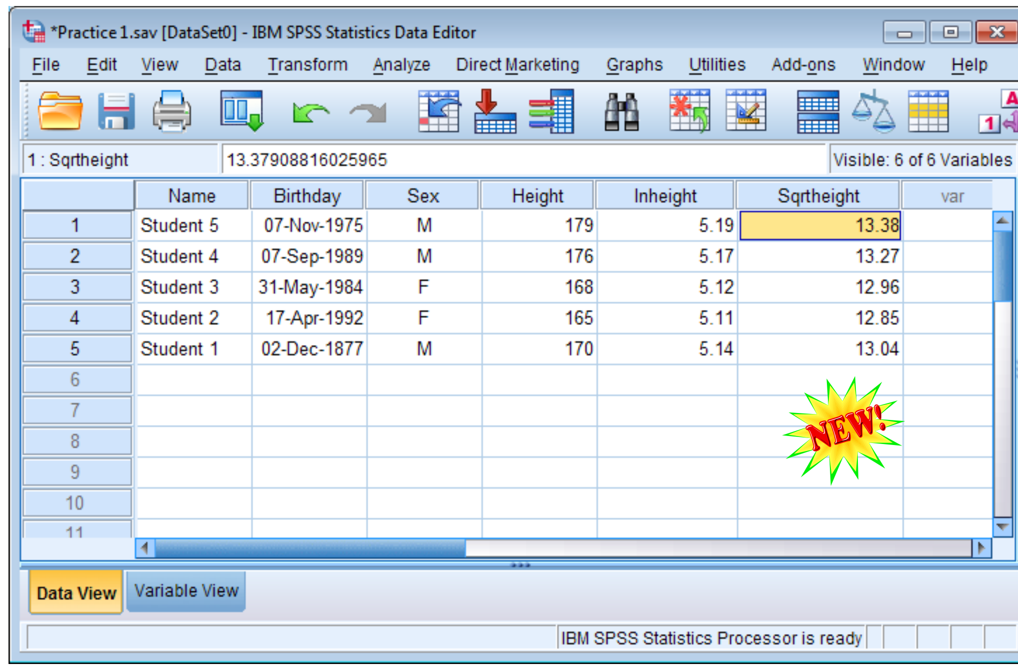


Practice 3

- Create a new variable named "sqrtheight" which is the square root of height.

Answer





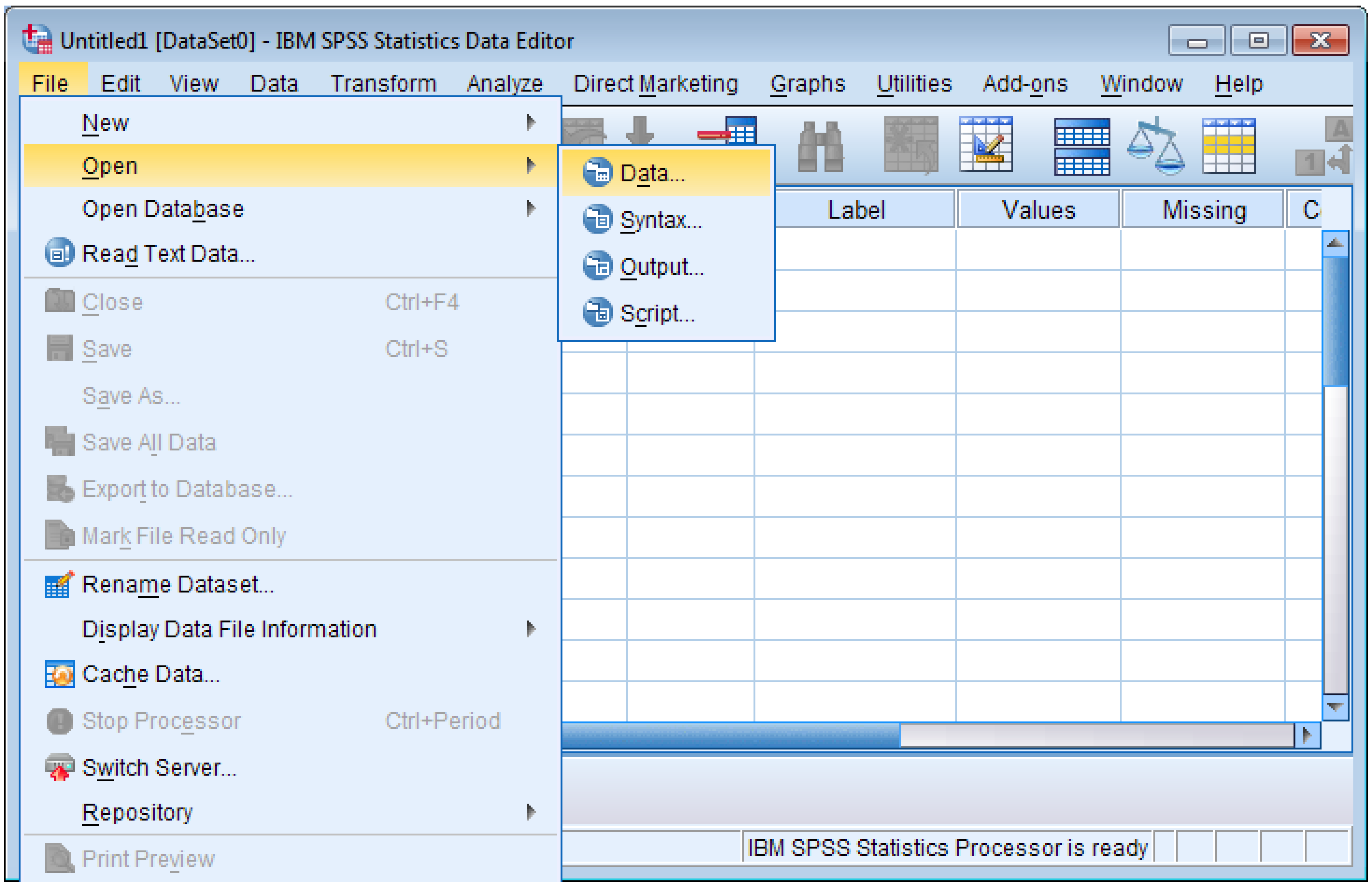
Basic analysis

Basic analysis of SPSS to be covered

- ❖ Frequencies
  - – This analysis produces frequency tables showing frequency counts and percentages of the values of individual variables.
- ❖ Descriptive
  - – This analysis shows the maximum, minimum, mean, and standard deviation of the variables
- ❖ Linear regression analysis
  - – Linear Regression estimates the coefficients of the linear equation

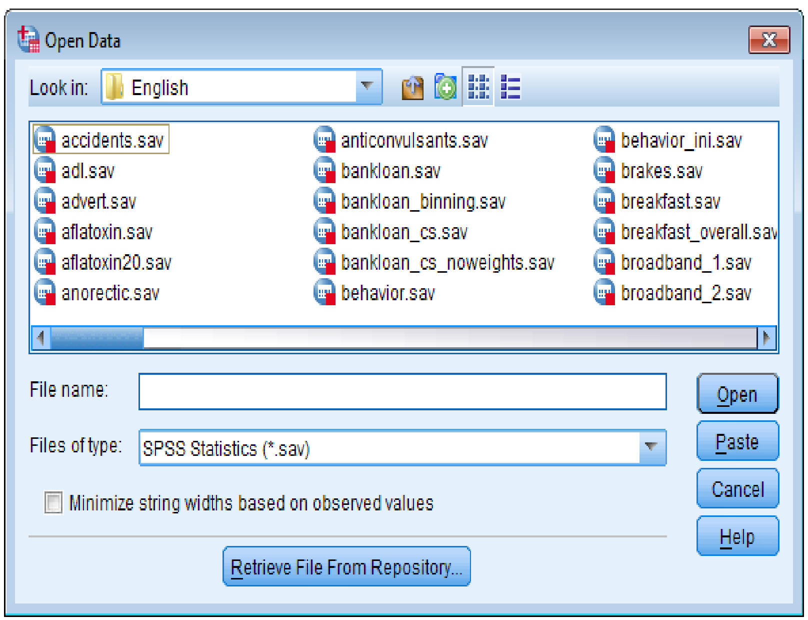
Opening the sample data

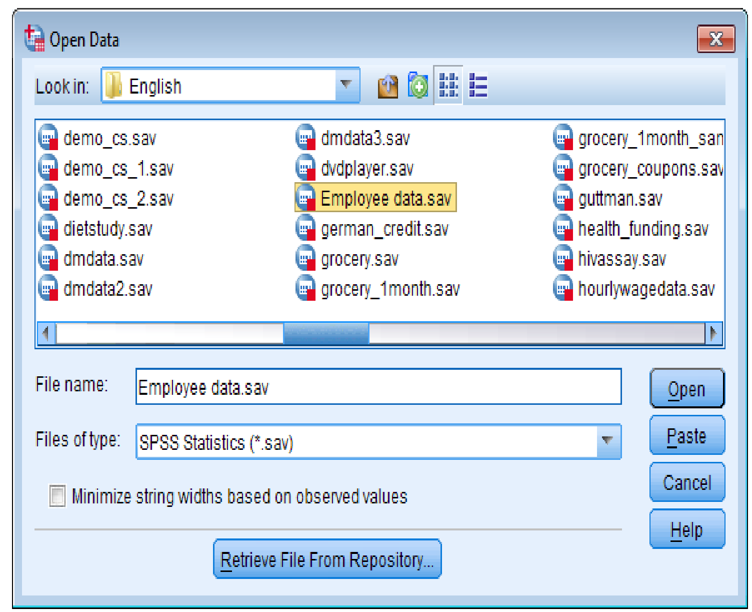
- Open 'Employee data.sav' from the SPSS
  - Go to "File," "Open," and Click Data



Opening the sample data

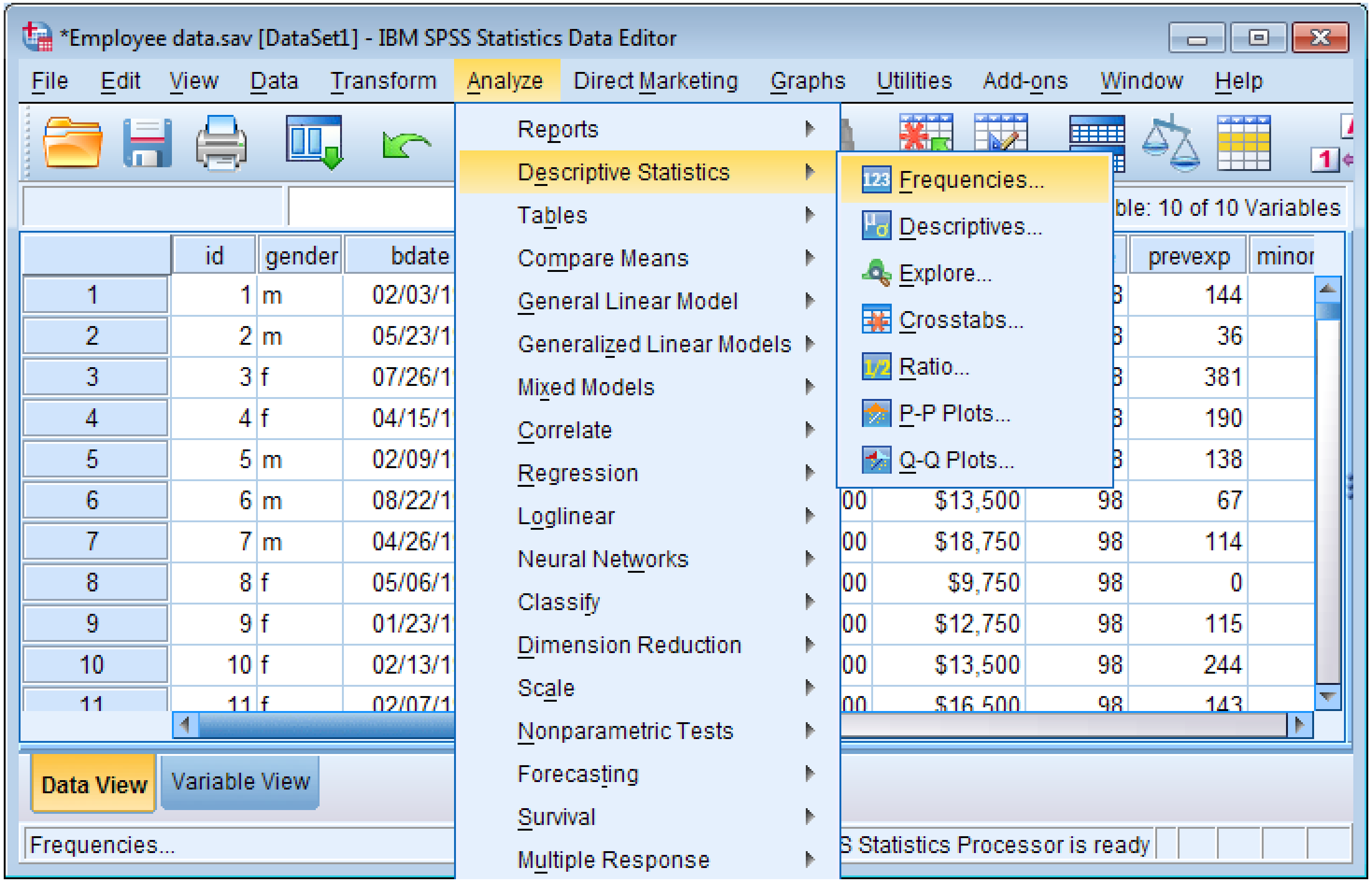
- Go to Program Files," "IBM," "SPSS," "Statistics," "20" and "Samples" folder. Then Select folder "English"
- Open "Employee Data.sav" file





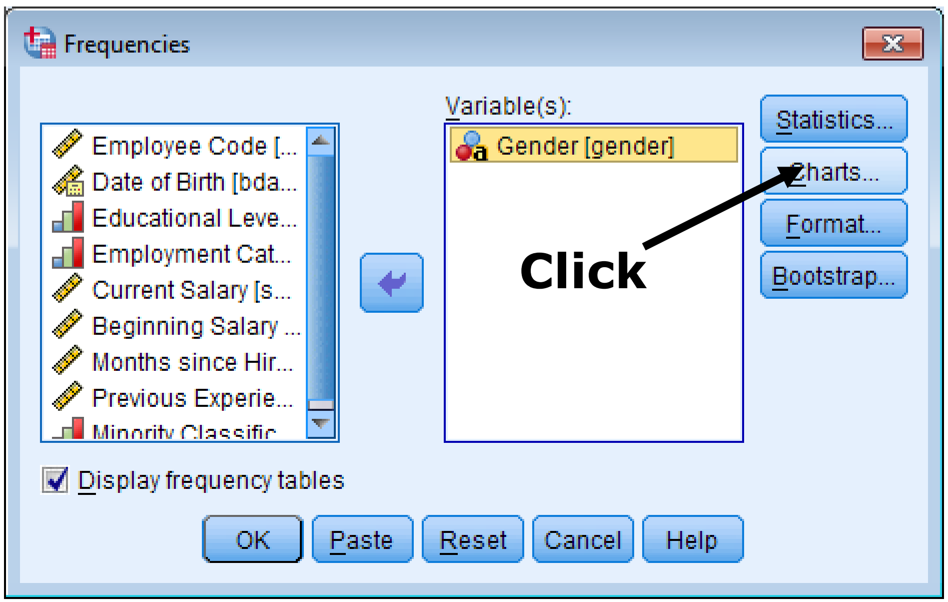
Frequencies

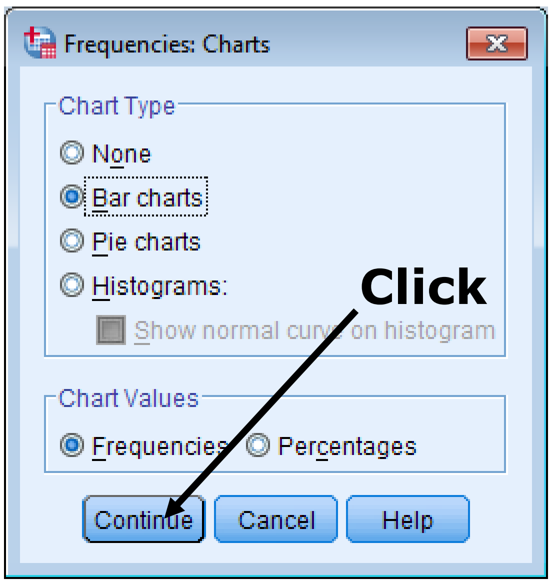
- Click ‘Analyze,’ ‘Descriptive statistics,’ then click ‘Frequencies’



Frequencies

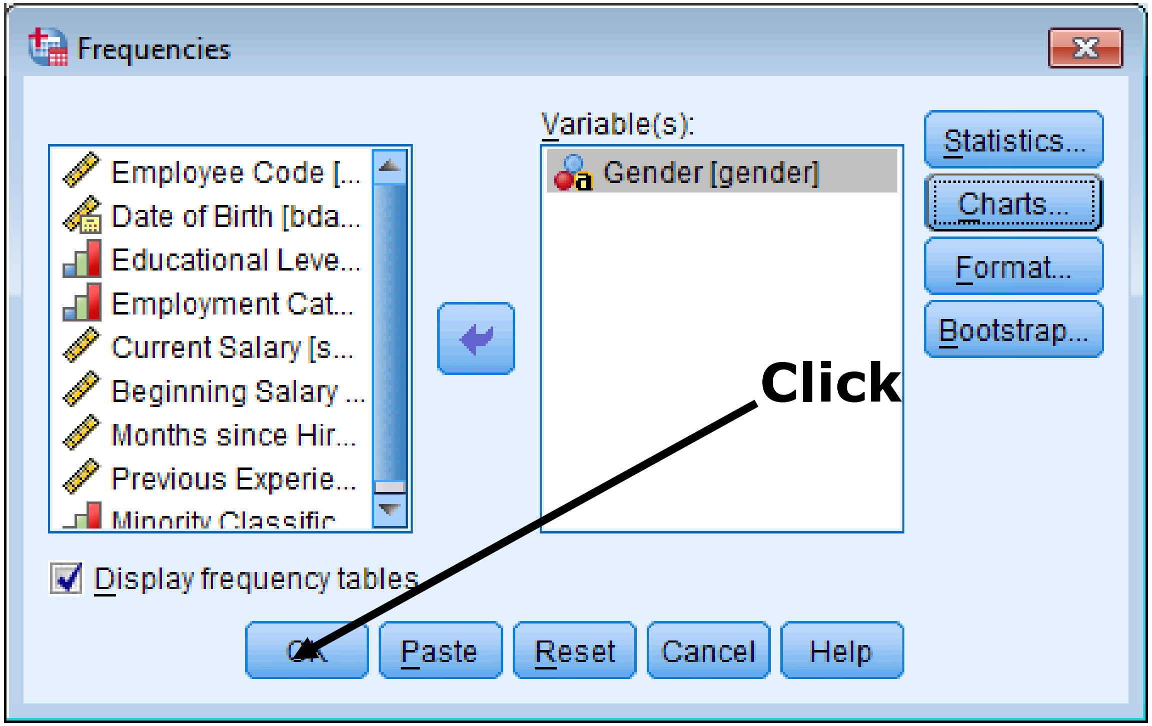
- Click gender and put it into the variable box.
- Click ‘Charts.’
- Then click ‘Bar charts’ and click ‘Continue.’

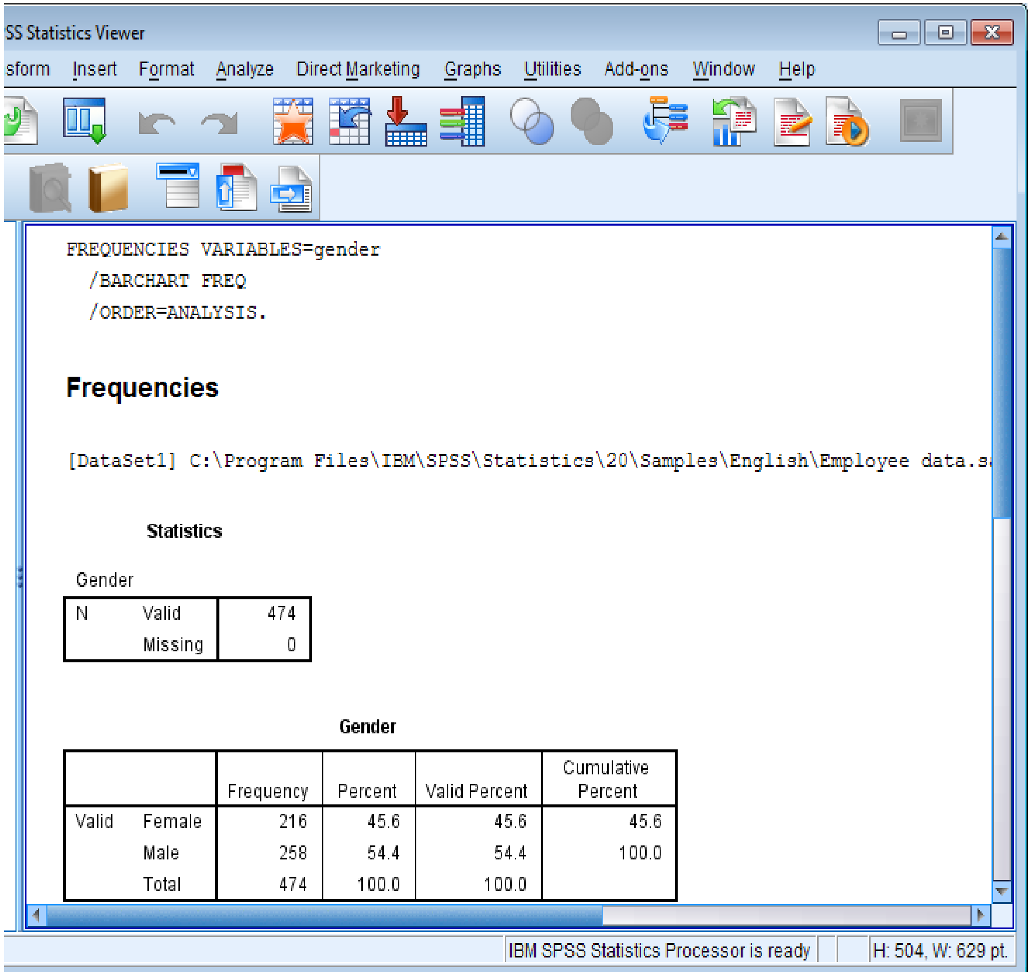




Frequencies

❖ Finally Click OK in the Frequencies box.






```
FREQUENCIES VARIABLES=gender
  /BARCHART FREQ
  /ORDER=ANALYSIS.
```

Frequencies

[DataSet1] C:\Program Files\IBM\SPSS\Statistics\20\Samples\English\Employee data.s

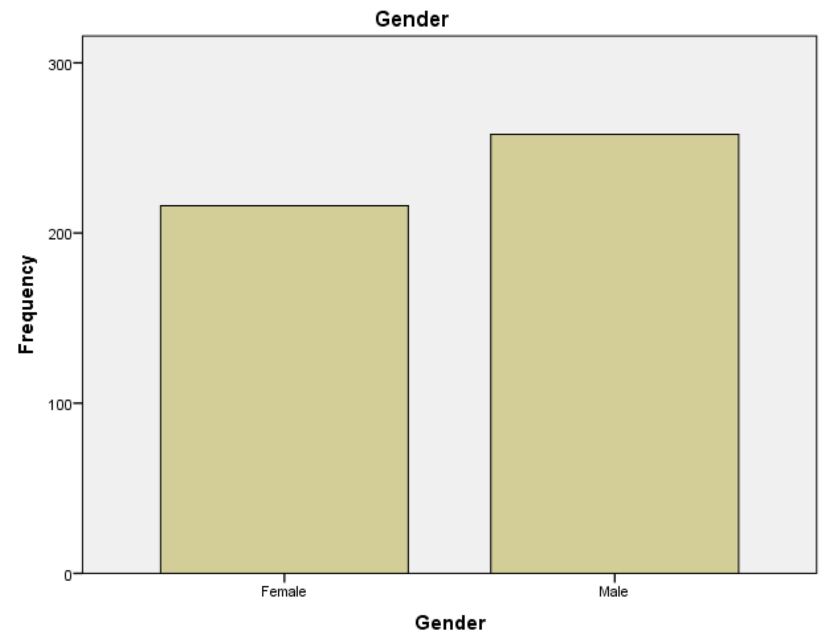
Statistics

Gender

N	Valid	474
	Missing	0

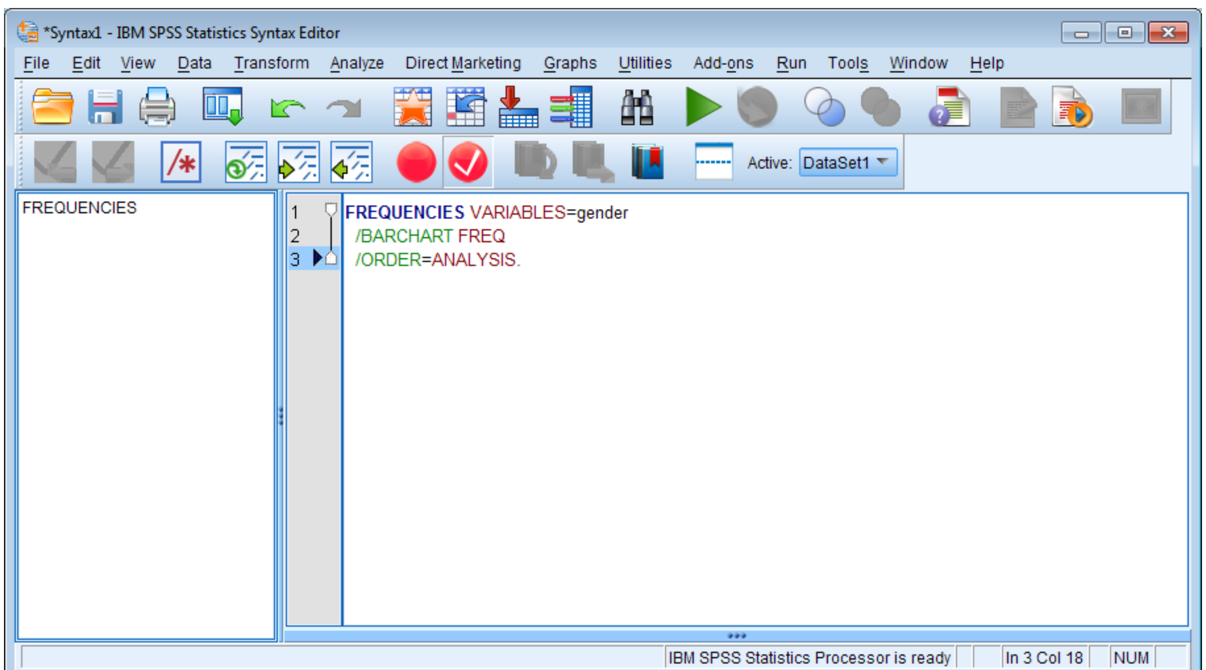
Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	216	45.6	45.6	45.6
	Male	258	54.4	54.4	100.0
	Total	474	100.0	100.0	



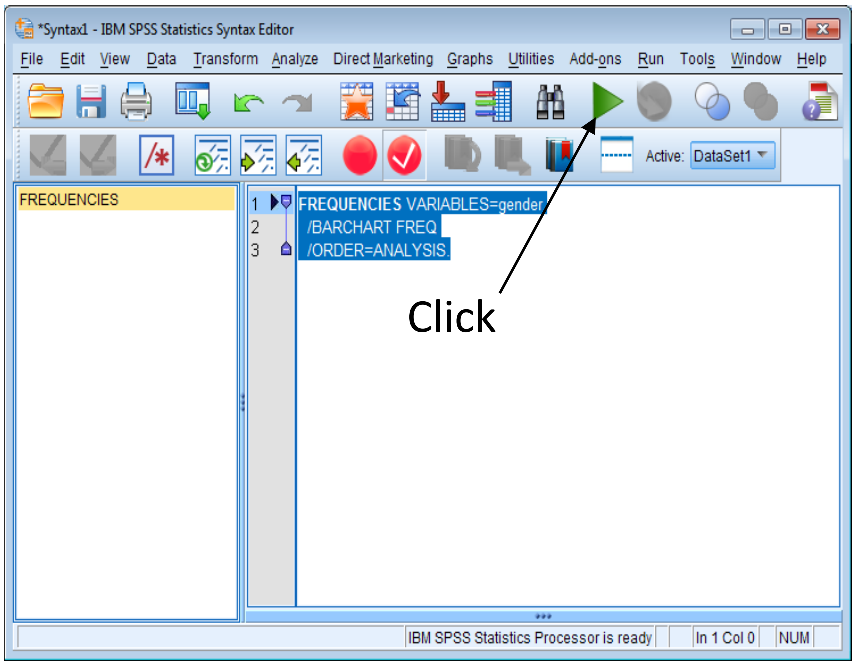
Using the Syntax editor

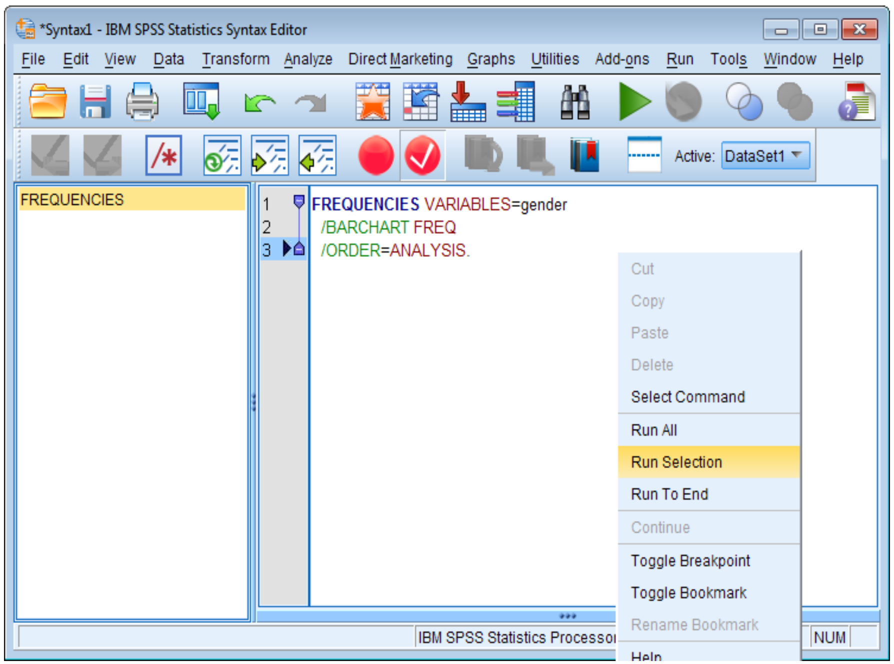
- Click ‘Analyze,’ ‘Descriptive statistics,’ then click ‘Frequencies.’
- Put ‘Gender’ in the Variable(s) box.
- Then click ‘Charts,’ ‘Bar charts,’ and click ‘Continue.’
- Click ‘Paste.’



Using the Syntax editor

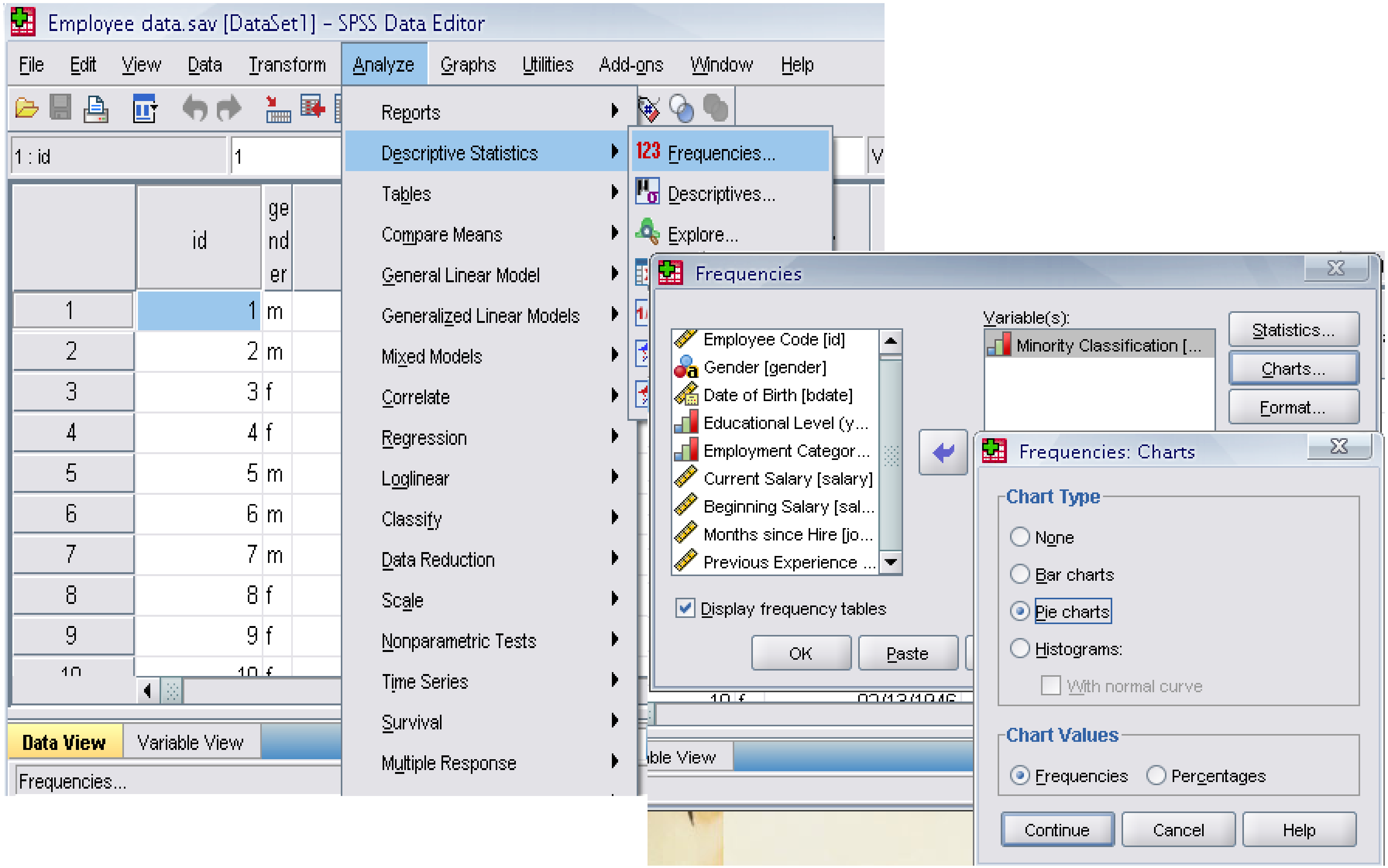
- Highlight the commands in the Syntax editor and then click the run icon.
- You can do the same thing by right clicking the highlighted area and then by clicking 'Run Current'



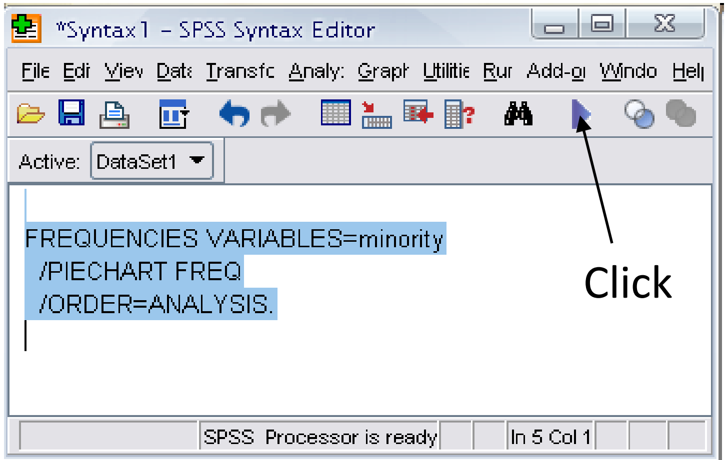


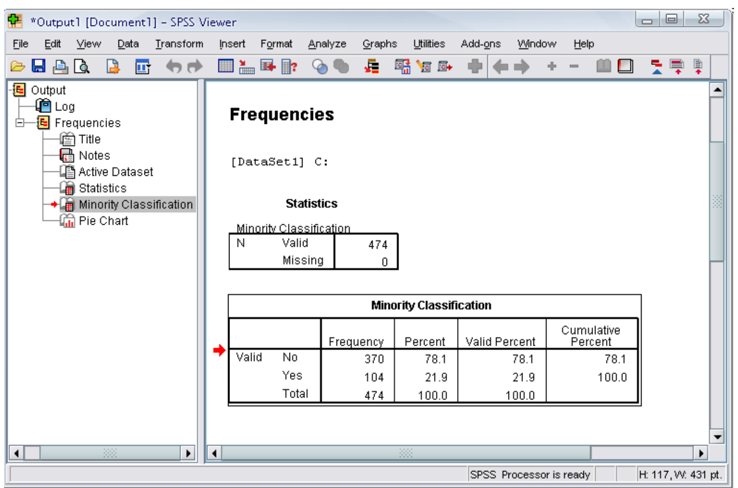
Practice 4

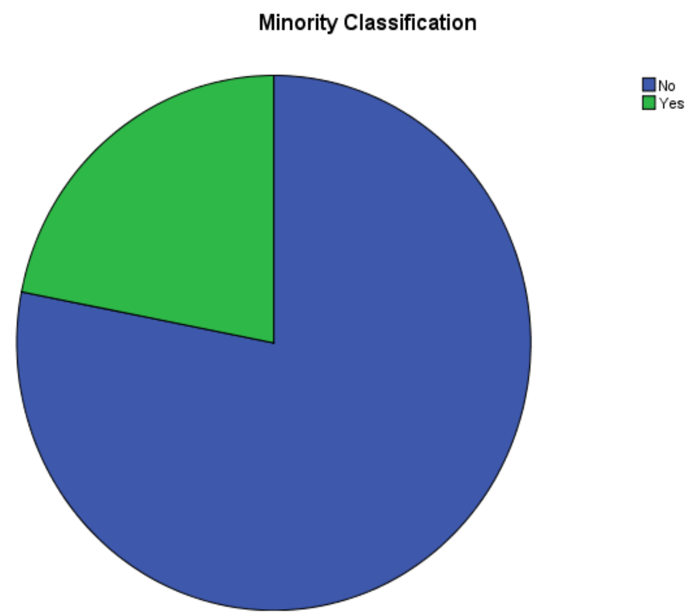
- ❖Do a frequency analysis on the variable “minority”
- ❖Create pie charts for it
- ❖Do the same analysis using the syntax editor


Employee data.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : id
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Frequencies...
Descriptives...
Explore...
Data View
Variable View
Frequencies
Variable(s):
Minority Classification [...
Statistics...
Charts...
Format...
Employee Code [id]
Gender [gender]
Date of Birth [bdate]
Educational Level (y...
Employment Categor...
Current Salary [salary]
Beginning Salary [sal...
Months since Hire [jo...
Previous Experience ...
Display frequency tables
OK
Paste
Frequencies: Charts
Chart Type
None
Bar charts
Pie charts
Histograms:
With normal curve
Chart Values
Frequencies
Percentages
Continue
Cancel
Help

Answer

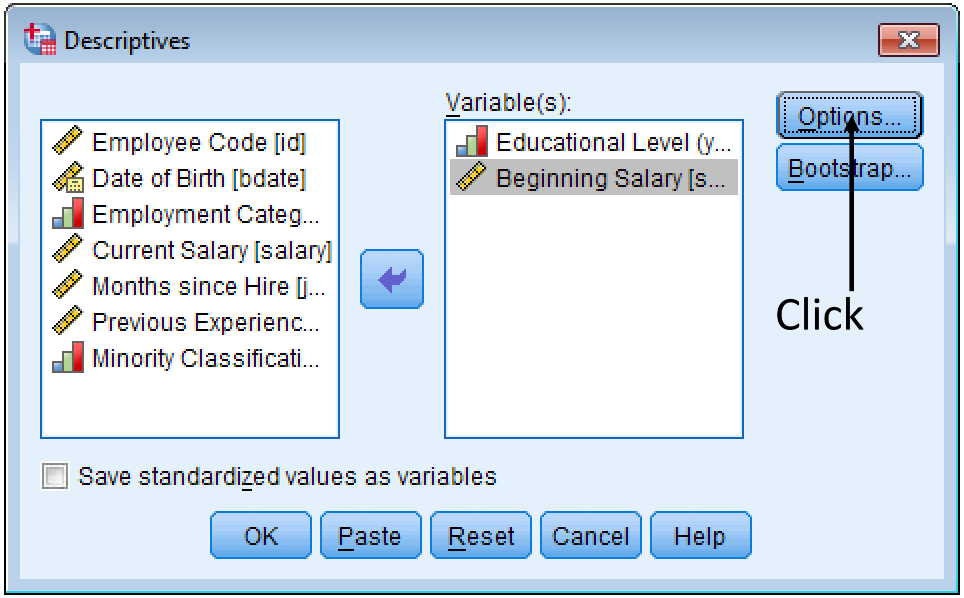

*Syntax1 - SPSS Syntax Editor
Active: DataSet1
FREQUENCIES VARIABLES=minority
/PIECHART FREQ
/ORDER=ANALYSIS.
Click
SPSS Processor is ready
In 5 Col 1


*Output1 [Document1] - SPSS Viewer
Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
Minority Classification
Pie Chart
Frequencies
[DataSet1] C:
Statistics
Minority Classification
N Valid 474
Missing 0
Minority Classification
Frequency Percent Valid Percent Cumulative Percent
Valid No 370 78.1 78.1 78.1
Yes 104 21.9 21.9 100.0
Total 474 100.0 100.0
SPSS Processor is ready
H: 117, W: 431 pt.


Minority Classification
No
Yes

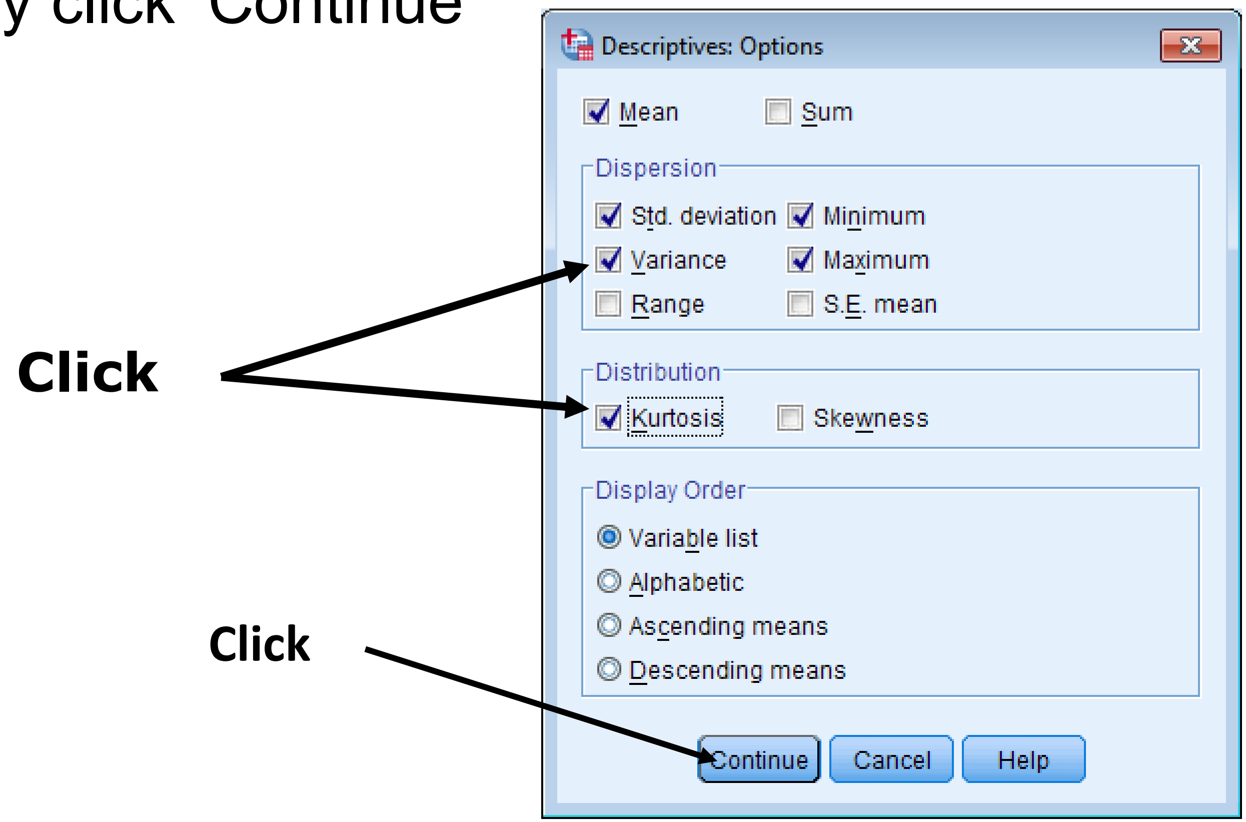
Descriptive

- ❖ Click ‘Analyze,’ ‘Descriptive statistics,’ then click ‘Descriptives…’
- ❖ Click ‘Educational level’ and ‘Beginning Salary,’ and put it into the variable box.
- ❖ Click Options



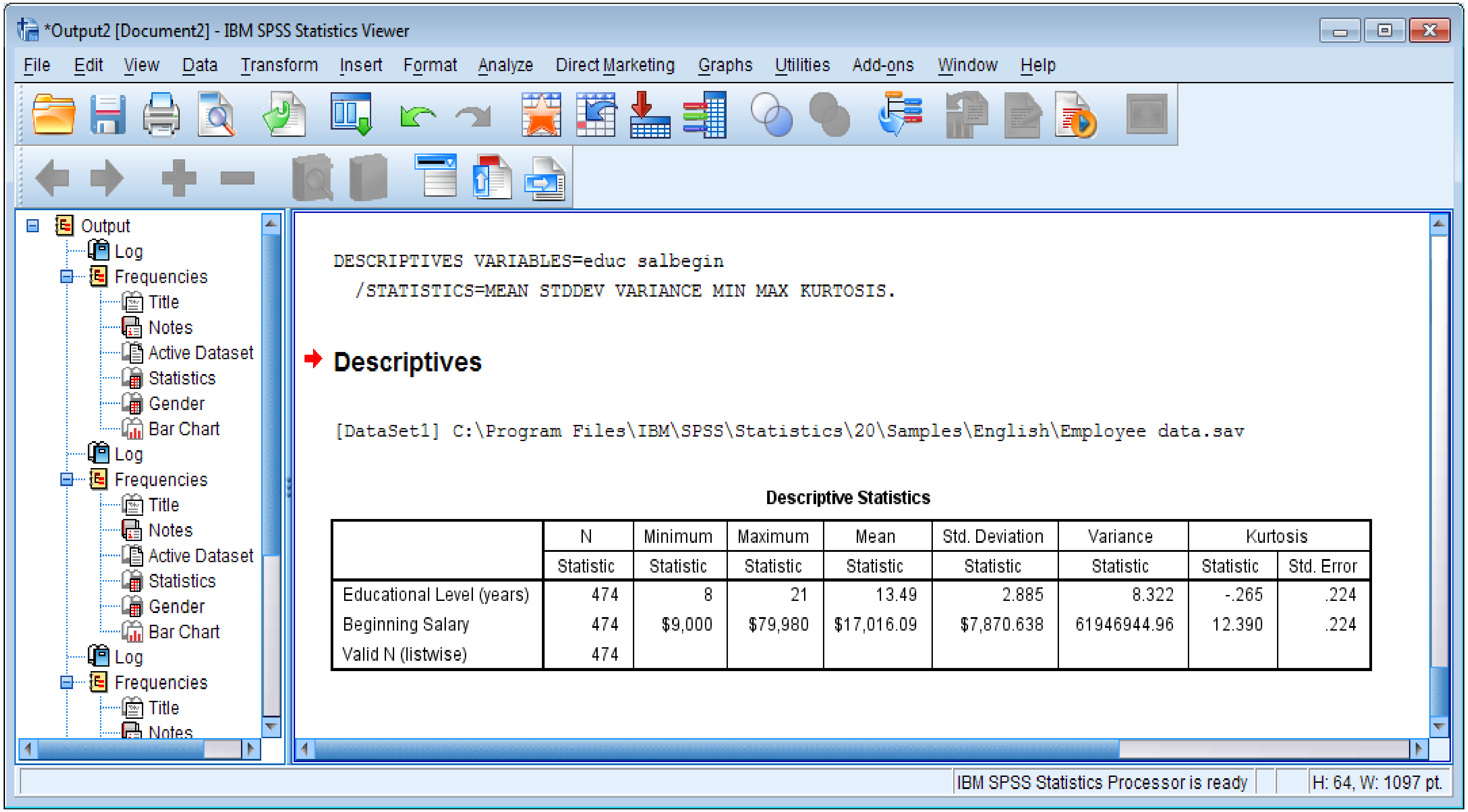
Descriptives

- The options allows you to analyze other descriptive statistics besides the mean and Std.
- Click ‘variance’ and ‘kurtosis’
- Finally click ‘Continue’



Descriptives

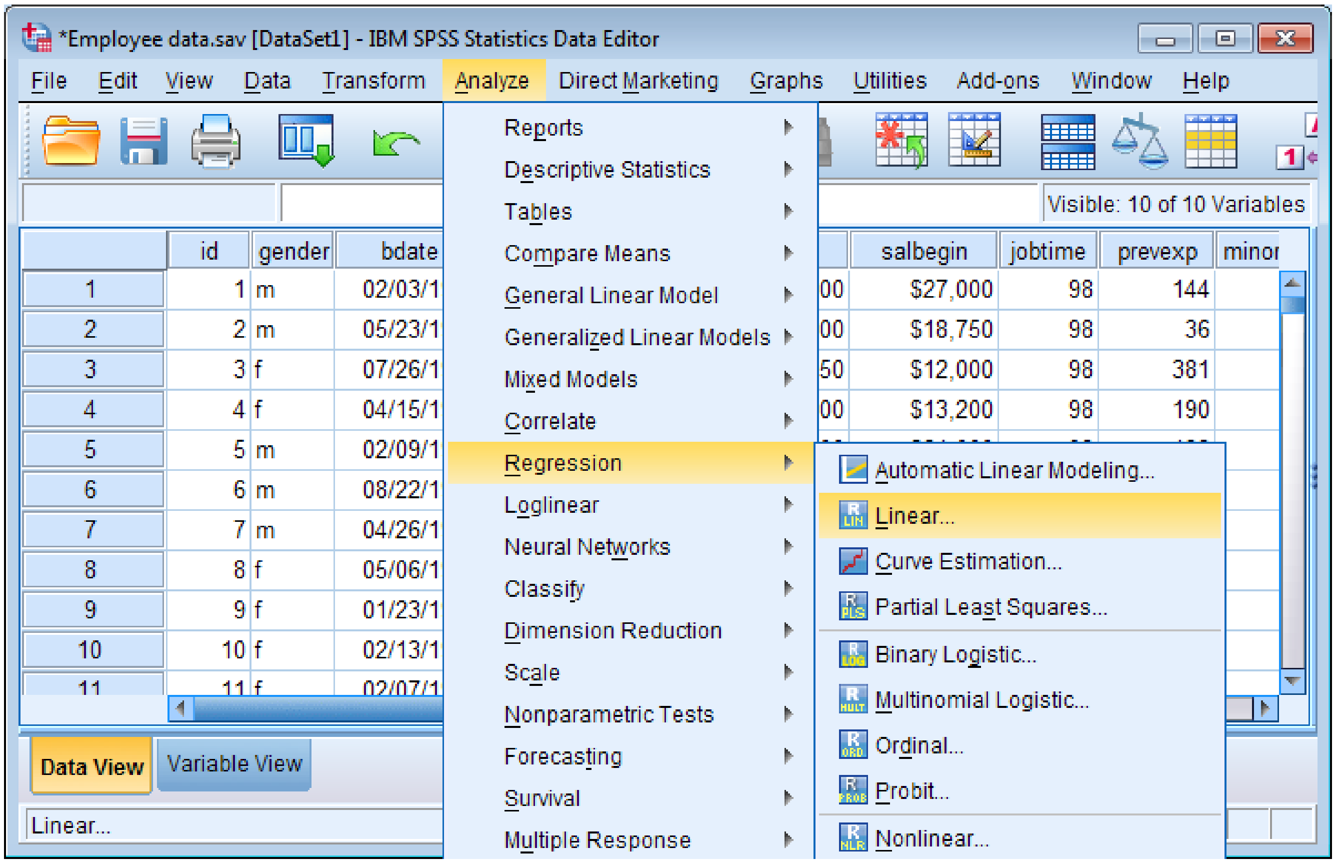
❖ Finally Click OK in the Descriptives box. You will be able to see the result of the analysis.



	N	Minimum	Maximum	Mean	Std. Deviation	Variance	Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Educational Level (years)	474	8	21	13.49	2.885	8.322	-.265	.224
Beginning Salary	474	$9,000	$79,980	$17,016.09	$7,870.638	61946944.96	12.390	.224
Valid N (listwise)	474							

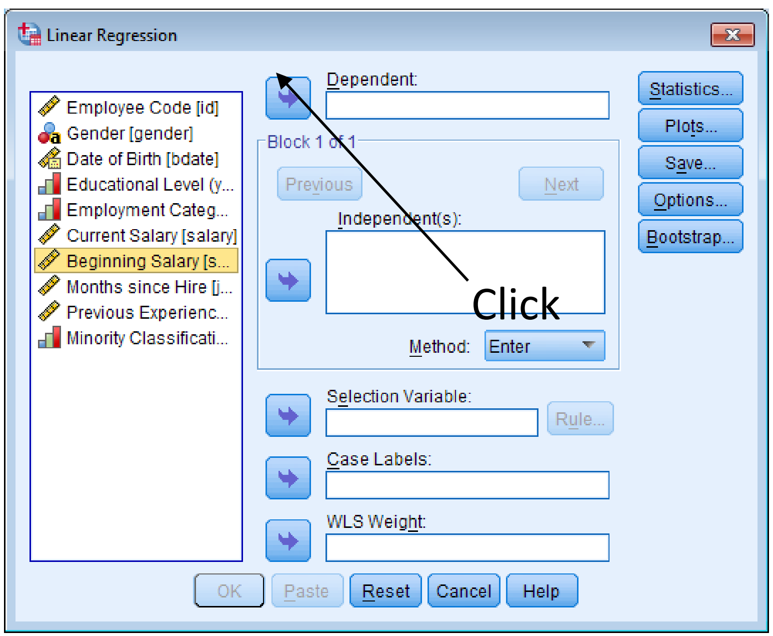
Regression Analysis

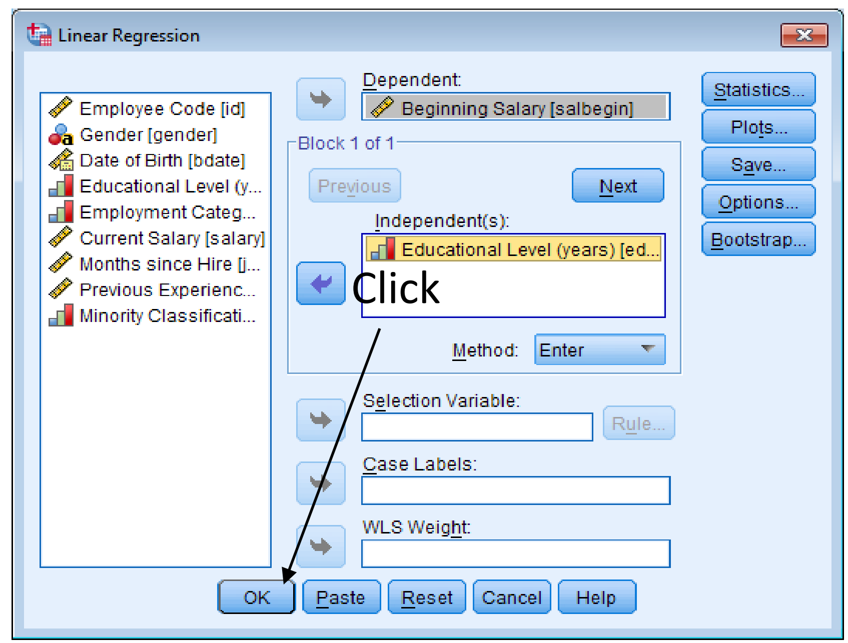
❖ Click 'Analyze,' 'Regression,' then click 'Linear' from the main menu.



Regression Analysis

- For example let's analyze the model $\text{salbegin} = \beta_0 + \beta_1 \text{edu} + \varepsilon$
- Put 'Beginning Salary' as Dependent and 'Educational Level' as Independent.





Regression Analysis

❖ Clicking OK gives the result

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.633[a]	.401	.400	$6,098.259

a. Predictors: (Constant), Educational Level (years)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11747808912	1	11747808912	315.897	.000[b]
	Residual	17553096053	472	37188762.82		
	Total	29300904965	473			

a. Dependent Variable: Beginning Salary

b. Predictors: (Constant), Educational Level (years)

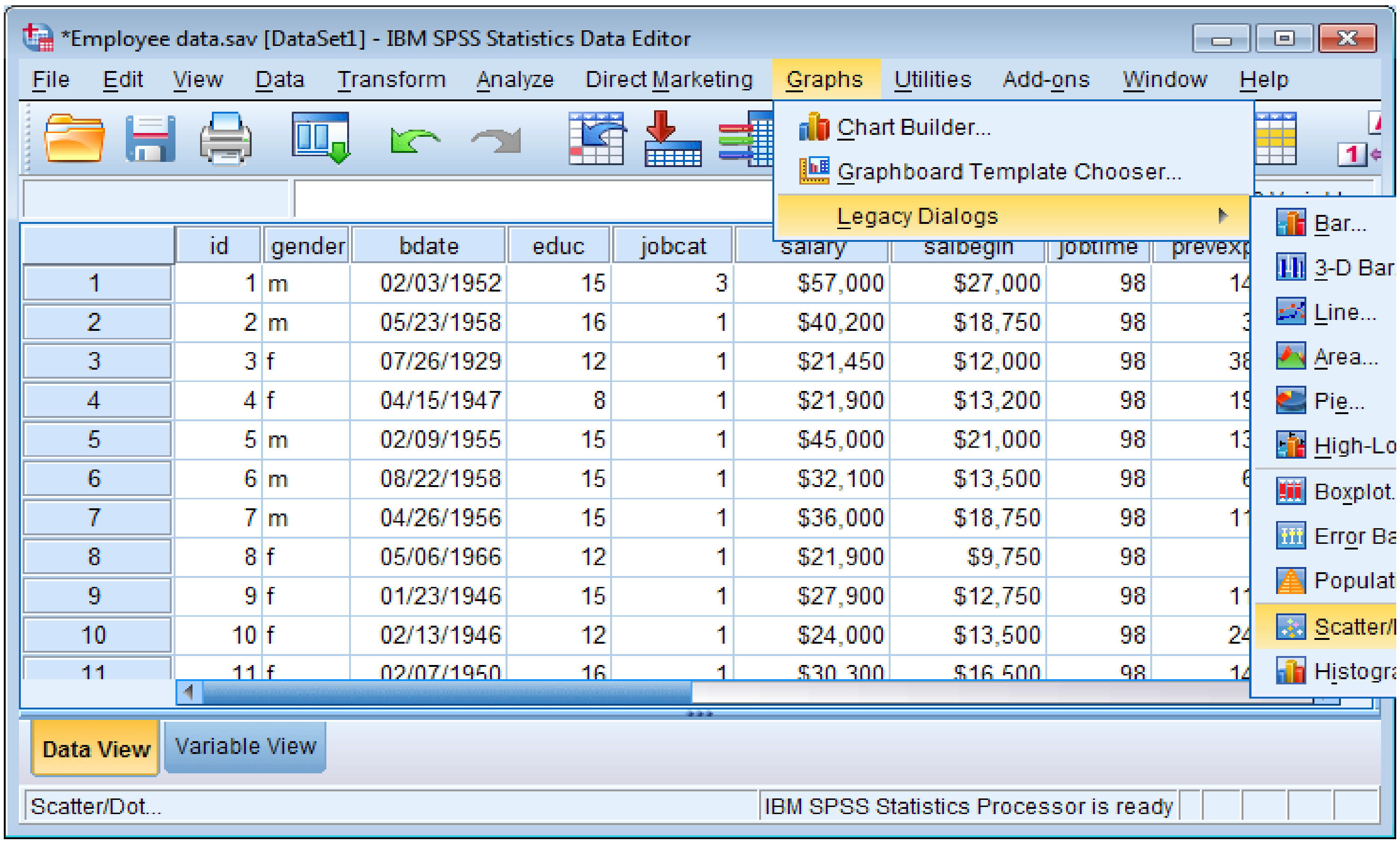
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6290.967	1340.920		-4.692	.000
	Educational Level (years)	1727.528	97.197	.633	17.773	.000

a. Dependent Variable: Beginning Salary

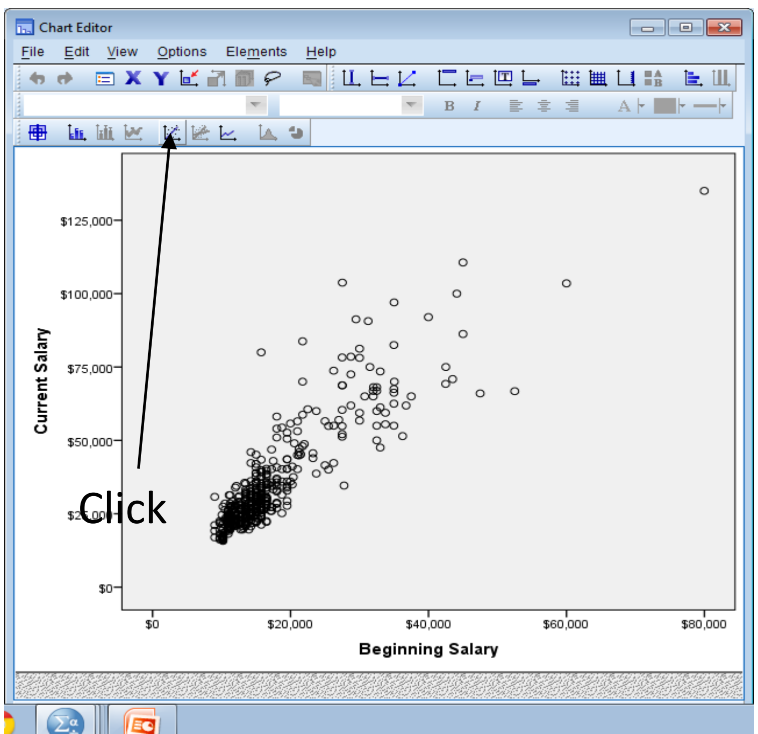
Plotting the regression line

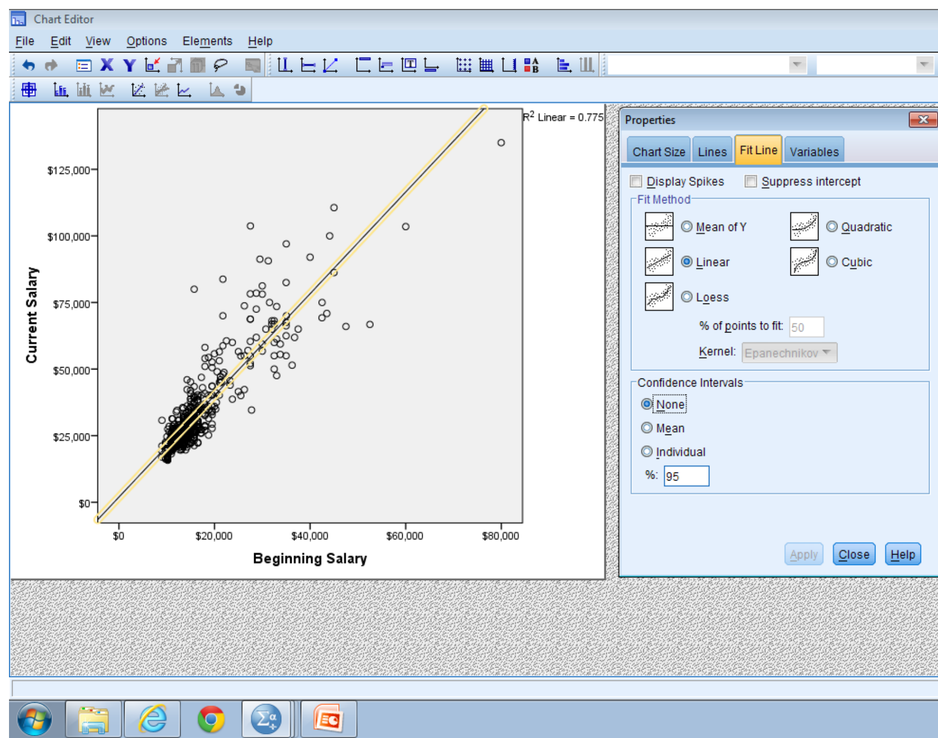
❖ Click ‘Graphs,’ ‘Legacy Dialogs,’ and ‘Scatter/Dot’ from the main menu.



Plotting the regression line

- Select Simple Scatter
- Set ‘Current Salary’ into the vertical axis box and ‘Beginning Salary’ in the horizontal axis box. Then click ‘OK’.
- Double click on the chart and open chart editor window
- Either click on “add fit line at total” action button or from menu select “Element” and then “fit line at total”





Practice 5

- Find out whether or not the previous experience of workers has any affect on their beginning salary?
  - Take the variable “salbegin,” and “prevexp” as dependent and independent variables respectively.

- Plot the regression line for the above analysis using the “scatter plot” menu.

THE END

Research Methods

Lecture 5-3: Discussion on Assignment 2

Agenda

- Introduction to overleaf
- Link to video resources available
- A small guide to start Overleaf
- Overleaf for Assignment 2

Introduction to overleaf

- **Latex is a software tool used for academic writings.**
- **It is of What You See Is What You Mean (WYSIWYM) type in contrary to most popular word processing packages like Word What You See Is What You Get (WYSIWYG).**
- **Why Overleaf instead of Latex**
  - Latex is a collection of software tools and need to install separately
  - Required many manual conversions
  - Identification of errors and their fixing is difficult in Latex
  - Overleaf is much user friendly and much robust to errors
  - Web based and conveniently stored in a cloud
  - Easy to track the errors in overleaf
  - Only clicking of one button generates the output on the right column window

Link to video lessons

- **Overleaf can be self learned using many video lessons available on the internet**
- **Few of such useful links are given below**
  - Video link on creating on Latex document: https://youtu.be/Qg2WtaSy-zQ
  - Video link on Sections and paragraph: https://youtu.be/3Ja6WZoXW3U
    Overleaf link: https://www.sharelatex.com/project/51e68895fdf7a9bd7b305993
  - Video link on Mathematics in Latex: https://youtu.be/2iEQoSIPD6o
    Overleaf link: https://www.sharelatex.com/project/51ed0480335f3d63700f3aa9
  - Video link on Images in Latex: https://youtu.be/21Y8LnKjfKE
    Overleaf link: https://www.sharelatex.com/project/51ee458d335f3d6370c9c58a
  - Video link on Bibliographies and natbib: https://youtu.be/8IMkTnR_KcA
    Overleaf link: https://www.sharelatex.com/project/51efbcd9335f3d637026853c
  - Video link on Tables and Matrices: https://youtu.be/1cEdqTMtHZQ
    Overleaf link: https://www.sharelatex.com/project/51f1324bc00460011608ea5c
  - Video link on longer documents: https://youtu.be/CxOCIfQhFmI
    Overleaf link: https://www.sharelatex.com/project/51f62ca03f7a0920417218de

Familiarization guide to overleaf

- A guide to familiarize with overleaf with the purpose of using it for assignment 2 is given below
- Link: Click here (filename: 5-4 How to use Overleaf)

Overleaf for Assignment 2

- **Find the zip file (file name: 5-5 Research Proposal.zip) containing the overleaf template, SLIIT logo, and sample research proposal generated.**
- **A guideline on writing the research proposal is included in the template itself.**
- **Students are strongly encouraged to use overleaf template in writing your research proposal for assignment 2.**
- **5% marks will be allocated for the effective use of the overleaf template in assignment 2**

Thank You..!

Choosing a Statistical Test

Choosing a Statistical test

- Let's ask Four Questions by ourselves

Q1: What type of data do you have?

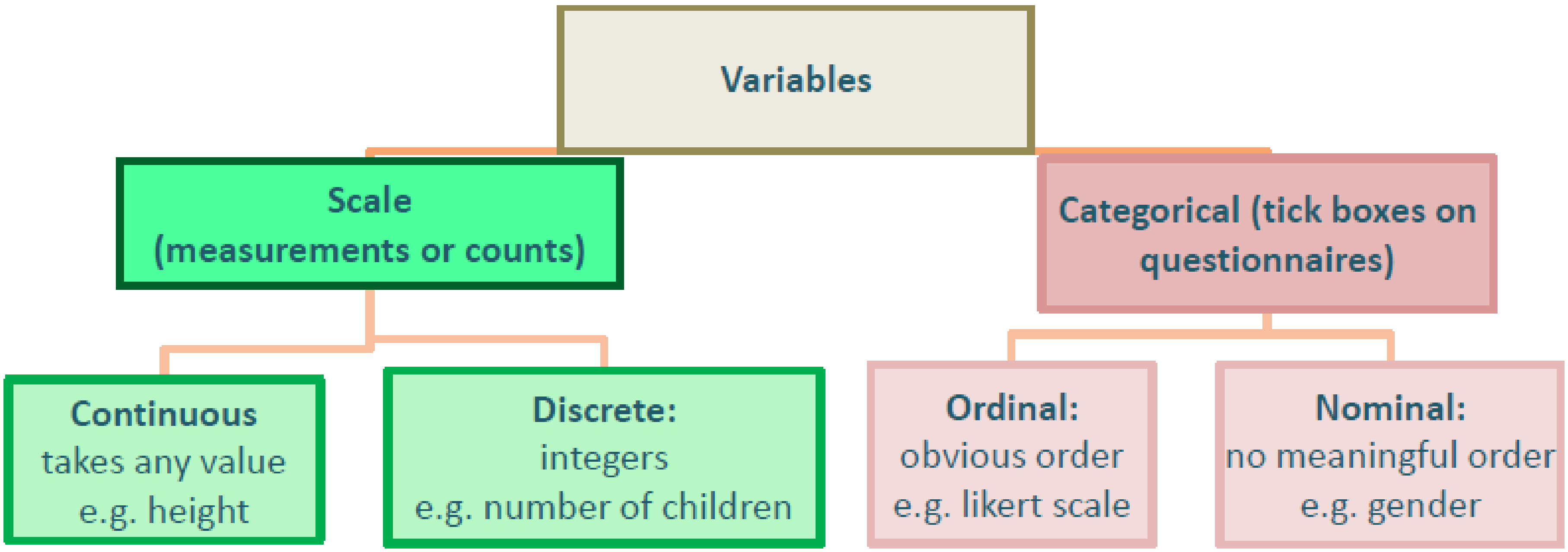
Categorical (often called rates)

- Mortality Rates (death/survival)
- Patient Falls Rates (fall/not fall)
- Compliance with Discharge Instructions (yes/no)

Quantitative

- Measurements
  - Temperatures
  - Blood Pressures
  - Pain Scale 1 to 10
- Counts
  - Number of Complications
  - Number of Hospital Visits
- Pre-established Scales

Categorization of variables



Choosing a Statistical test

- Let’s ask Four Questions by ourselves

Q2: How many samples do you have?

- One Sample
  - Usually means I am making a comparison against a historic or global value
- Two Samples
  - Traditional control group vs. test group in an experiment
  - Comparing the one group to the other group
- Two Samples Special (or Two-ish Samples)
  - One sample of people, but two different measurements made
  - Two samples of people, but for each person in one sample, there is a natural pairing with one person in the other sample
- Three or More Samples
  - Comparing the three groups to each other

Choosing a Statistical test

- Let's ask Four Questions by ourselves

Q3: What is the test supposed to do?

Compare the data

- Does my one sample match or differ from a global or historic value?
- Are my two samples the same or different from each other?
- Are my three or more samples the same or different from each other?

Seek a relationship

- Note: This is only for the two samples – special
- Does one measurement predict the other?
- Does one measurement depend on the other?
- Is there a trend between the two sets of measurements?

What is the research question? Are we asking "is it this," so that we need to conduct a hypothesis test? Or are we asking "what is it," so that we need to calculate a point estimate or a confidence interval?

Choosing a Statistical test

- Let's ask Four Questions by ourselves

Q4: What assumptions can we safely make about the data?

- Can we assume that the data are normally distributed?
- Can we assume the variances of two populations are equal?
- Are the groups dependent or independent?

Data Analysis: Common Requirements

Get a description of a group of data (e.g. get central tendency)

Compare one group of data to a hypothetical value (e.g. mean to a certain value)

Compare two groups (e.g. means)

Compare three or more groups

Find association between two variables

Make a prediction

Data Analysis: Common Requirements and SOLUTIONS

Get a description of a group of data (e.g. get central tendency)

- Ratio/Interval data + Normally distributed: Mean, SD
- Ratio/Interval data + Non-normally distributed: Median, Interquartile range
- Ordinal data : Median, Interquartile range
- Nominal data : Proportion

Data Analysis: Common Requirements and SOLUTIONS

Compare one group of data to a hypothetical value

- Ratio/Interval data + Normally distributed: One Sample t test
- Ratio/Interval data + Non-normally distributed: Wilcoxon test
- Ordinal data : Wilcoxon test
- Nominal data : Chi-square or Binomial test

Data Analysis: Common Requirements and SOLUTIONS

Comparison of two groups

- Unpaired + Ratio/Interval data + Normally distributed: Unpaired t test
- Unpaired + Ratio/Interval data + Non-normally distributed: Mann-Whitney test
- Unpaired + Ordinal data : Mann-Whitney test
- Unpaired + Nominal data : Fisher’s Test, Chi-square for large samples
- Paired + Ratio/Interval data + Normally distributed: Paired t test
- Paired + Ratio/Interval data + Non-normally distributed: Wilcoxon test
- Paired + Ordinal data : Wilcoxon test
- Paired + Nominal data : McNemar's test

Data Analysis: Common Requirements and SOLUTIONS

Compare three or more groups

- Unmatched + Ratio/Interval data + Normally distributed: One way ANOVA
- Unmatched + Ratio/Interval data + Non-normally distributed: Kruskal-Wallis test
- Unmatched + Ordinal data : Kruskal-Wallis test
- Unmatched + Nominal data : Chi-square test
- Matched + Ratio/Interval data + Normally distributed: Repeated measures ANOVA
- Matched + Ratio/Interval data + Non-normally distributed: Friedman test
- Matched + Ordinal data : Friedman test
- Matched + Nominal data : Cochrane Q test

Ref: https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php
https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php
https://statistics.laerd.com/spss-tutorials/one-way-anova-repeated-measures-using-spss-statistics.php
https://statistics.laerd.com/spss-tutorials/cochrans-q-test-in-spss-statistics.php

Data Analysis: Common Requirements and SOLUTIONS

Measure association between two variables

- Ratio/Interval data + Normally distributed: Pearson correlation
- Ratio/Interval data + Non-normally distributed: Spearman correlation
- Ordinal data : Spearman correlation
- Nominal data : Contingency coefficients

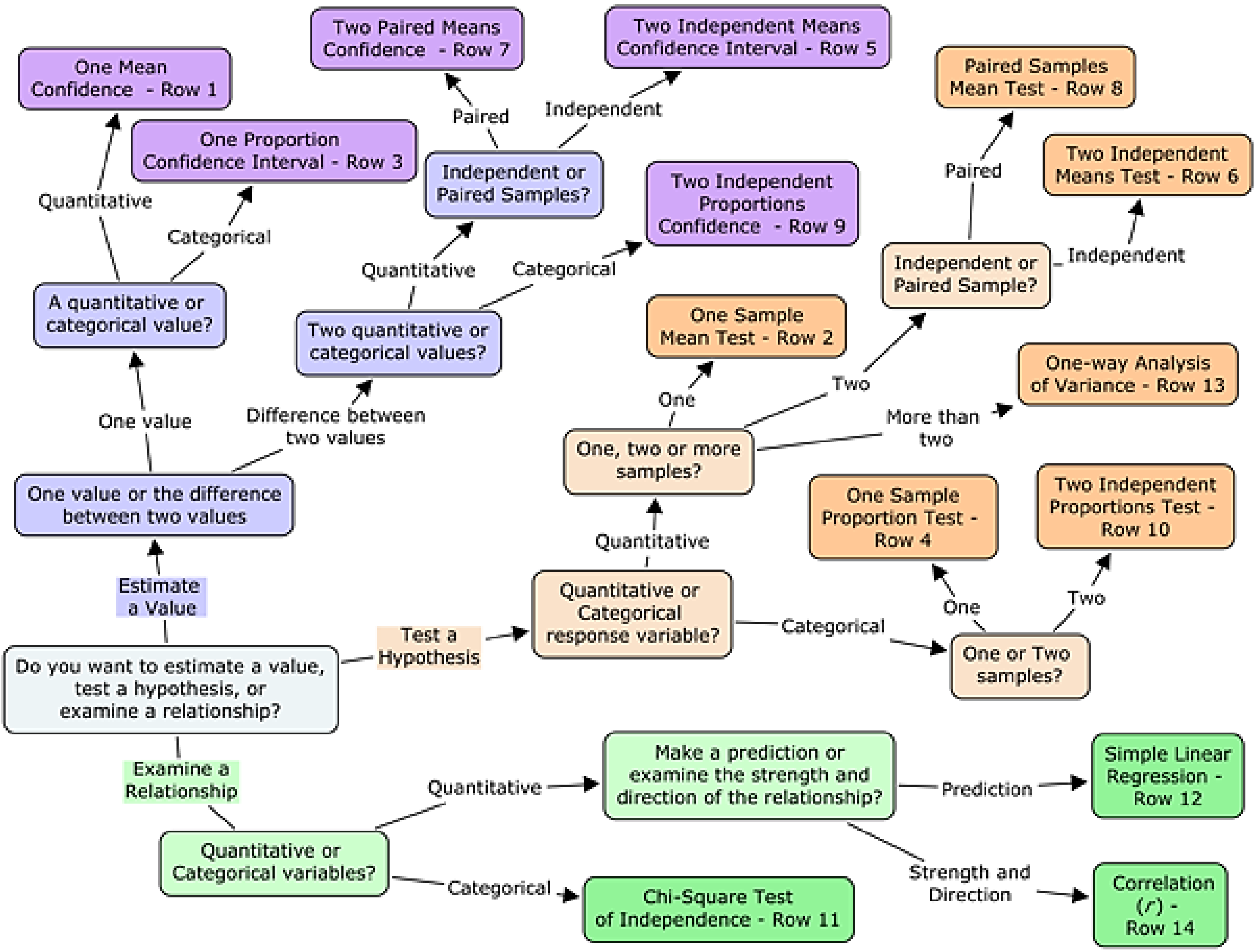
Data Analysis: Common Requirements and SOLUTIONS

Find a prediction

- From one variable + Ratio/Interval data + Normally distributed: Simple linear regression
- From one variable + Ratio/Interval data + Non-normally distributed: Nonparametric regression
- From one variable + Ordinal data : Nonparametric regression
- From one variable + Nominal data : Simple logistic regression
- From several variables + Ratio/Interval data + Normally distributed: Multiple linear regressions or Multiple nonlinear
- From several variables + Ratio/Interval data + Non-normally distributed: Multiple linear regressions or Multiple nonlinear
- From several variables + Ordinal data : Multiple linear regressions or Multiple nonlinear
- From several variables + Nominal data : Multiple logistic regressions

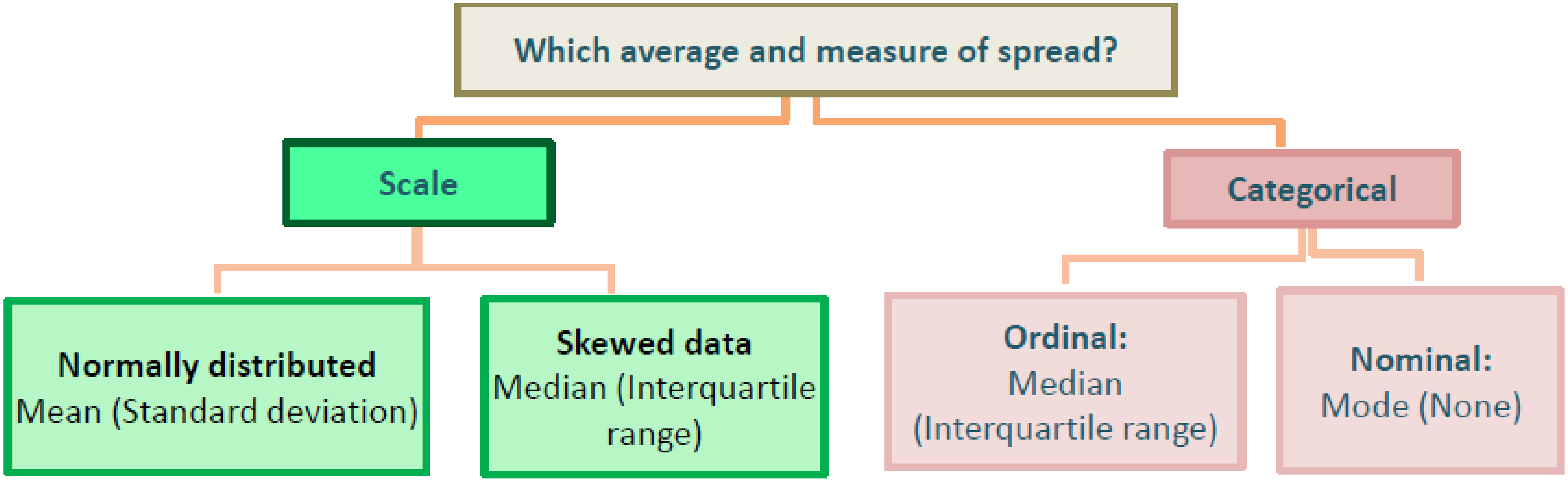
Ref: https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php
https://statistics.laerd.com/spss-tutorials/binomial-logistic-regression-using-spss-statistics.php
https://statistics.laerd.com/spss-tutorials/multinomial-logistic-regression-using-spss-statistics.php

Choosing a Statistical test



https://onlinecourses.science.psu.edu/stat500/node/67/

Summary statistics

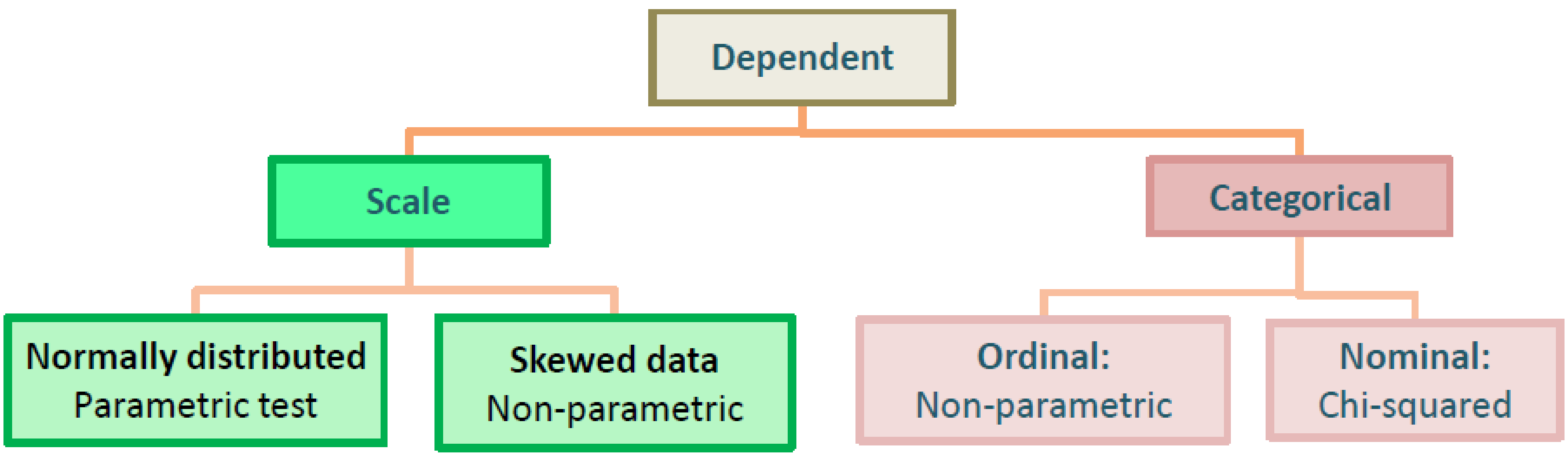


Summary of descriptive and graphical statistics

Chart	Variable type	Purpose	Summary Statistics
Pie Chart or bar chart	One Categorical	Shows frequencies/ proportions/percentages	Class percentages
Stacked / multiple bar	Two categorical	Compares proportions within groups	Percentages within groups
Histogram	One scale	Shows distribution of results	Mean and Standard deviation
Scatter graph	Two scale	Shows relationship between two variables and helps detect outliers	Correlation co-efficient
Boxplot	One scale/ one categorical	Compares spread of values	Median and IQR
Line Chart	Scale by time	Displays changes over time Comparison of groups	Means by time point
Means plot	One scale/ 2 categorical	Looks at combined effect of two categorical variables on the mean of one scale variable	Means

Appropriate statistical methods for research

- Which is the dependent (outcome) variable and what type of variable is it?
- Which are the independent (explanatory) variables, how many are there and what data types are they?



Common Single Comparison Tests

Comparing:	Dependent (outcome) variable	Independent (explanatory) variable	Parametric test (data is normally distributed)	Non-parametric test (ordinal/ skewed data)
The averages of two INDEPENDENT groups	Scale	Nominal (Binary)	Independent t-test	Mann-Whitney test/ Wilcoxon rank sum
The averages of 3+ independent groups	Scale	Nominal	One-way ANOVA	Kruskal-Wallis test
The average difference between paired (matched) samples e.g. weight before and after a diet	Scale	Time/ Condition variable	Paired t-test	Wilcoxon signed rank test
The 3+ measurements on the same subject	Scale	Time/ condition variable	Repeated measures ANOVA	Friedman test

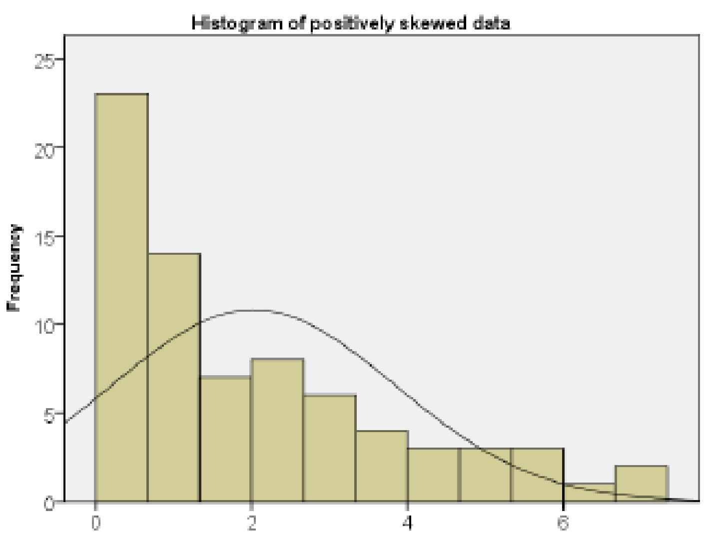
Tests of association

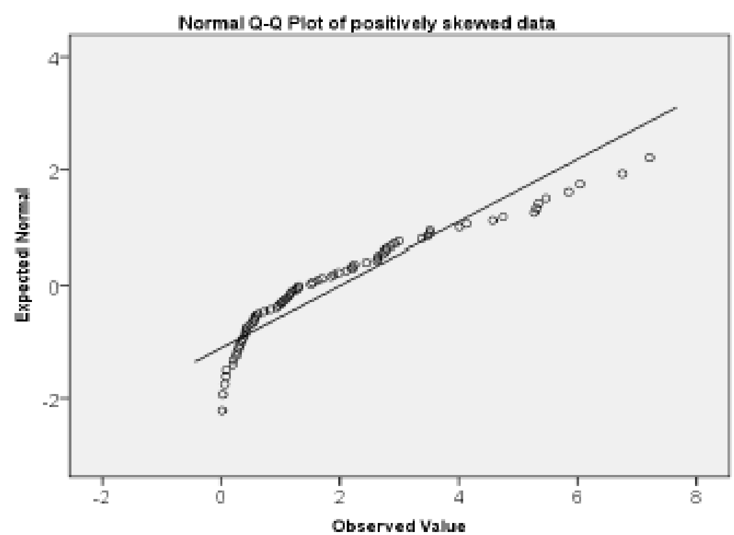
Comparing:	Dependent (outcome) variable	Independent (explanatory) variable	Parametric test (data is normally distributed)	Non-parametric test (ordinal/ skewed data)
Relationship between 2 continuous variables	Scale	Scale	Pearson's Correlation Coefficient	Spearman's Correlation Coefficient
Predicting the value of one variable from the value of a predictor variable or looking for significant relationships	Scale	Any	Simple Linear Regression	Transform the data
	Nominal (Binary)	Any	Logistic regression	
Assessing the relationship between two categorical variables	Categorical	Categorical		Chi-squared test

Assumption of normality

- Plotting a histogram or QQ plot of the variable of interest will give an indication of the shape of the distribution.
- Histograms should peak in the middle and be approximately symmetrical about the mean.
- If data is normally distributed, the points in QQ plots will be close to the line.

Below are some examples of very skewed data (i.e. non-normal).





Statistical tests for normality

- *Shapiro-Wilk and Kolmogorov-Smirnoff tests*
    - *but for small sample sizes (n < 20), the tests are unlikely to detect non-normality and for larger sample sizes (n > 50), the tests can be too sensitive.*
- *They are also sensitive to outliers so use histograms (large samples) or QQ plots (small samples).*

Parametric test	What to check for normality
Independent t-test	Dependent variable by group
Paired t-test	Paired differences
One-way ANOVA	Residuals
Repeated measures ANOVA	Residuals at each time point
Pearson's correlation coefficient	Both variables are normally distributed
Simple linear regression	Residuals

Assumption of Homogeneity of variances

- Test: ***Levene's test***
- *Use: Used to test the equality of variances when comparing the means of independent groups*
  - *e.g. Independent t-tests and ANOVA*
- **Note:** The violation of this assumption is more serious than violation of the assumption of normality
  - But both t-tests and ANOVA are fairly robust to deviations from this assumption
- Interpretation:
  - If the p-value is less than 0.05 reject H_0 and conclude that the assumption of equal variances has not been met.

Assumption of Sphericity

- Test: ***Mauchly's test***
- *Use: Tests for sphericity - a measure of whether variances of the differences between all repeated measures are all equal.*
- *Note: If the assumption is not met, the F-statistic is positively biased leading to an increased risk of a type 1 error.*
- ***Interpretation:***
  - Significant when p-value < 0.05 meaning there are significant differences between the variance of differences, i.e. condition of sphericity is not met.
  - If the assumption is not met, use the Greenhouse-Geisser correction to the degrees of freedom which appears in the standard output.

Assumption of independent observations

- For most tests, it is assumed that the observations are independent.
- The results for one subject (unit of interest which could be a person, an observation, a day etc.) are not affected by another.
- Examples of data which is not independent are
  - Repeated measures on the same subject and observations over time
    - check the Durbin Watson test for regression
  - when subjects are nested within groups with a common influence
    - e.g. children within classes who may be influenced by the teacher

Independent t-test

- Dependent variable: Continuous
- Independent variable: Binary (Group)
- ***Use:*** A t-test is used to compare the means of two independent groups. Independent groups means that different people are in each group.
- ***Plot:*** Box-plots (exploratory)or Confidence Interval plots with results

Assumptions	**How to check**	**What to do if assumption is not met**
Normality: dependent variables should be normally distributed within each group	Histograms of dependent variables per group / Shapiro Wilk	Mann-Whitney / Wilcoxon rank sum
Homogeneity of variance	Levene's test * (part of standard SPSS output)	Use bottom row of t-test output in SPSS

- *Levenes test: If the assumption of homogeneity is not met,
  - not using the pooled estimate for the error term for the t-statistic and
  - making adjustments to the degrees of freedom using the Welch-Satterthwaite method.
  - SPSS does the automatically in the "Equal variances not assumed" row.
  - Alternatively, a Mann-Whitney test can be carried out.
- ***Interpretation:***
  - If the p-value < 0.05, there is a significant difference between the means of the two groups. Report the means of the two groups or the mean difference and confidence interval from the SPSS output to describe the difference.
- *SPSS: Analyze →Compare means → Independent-samples T-test*

Mann-Whitney test

- The Mann-Whitney test is also known as the Mann–Whitney–Wilcoxon (MWW), Wilcoxon rank-sum test, or Wilcoxon–Mann–Whitney test.
- Non-parametric equivalent to the independent t-test
- Dependent variable: Ordinal/ Continuous
- Independent variable: Binary(Group)
- ***Use:***
  - It is used to compare whether two groups containing different people are the same or not. The Mann-Whitney test ranks all of the data and then compares the sum of the ranks for each group to determine whether the groups are the same or not.
  - If the distribution of scores for both groups have the same shape, the medians can be compared. If not, use the default test which compares the mean ranks.
- ***Plot:*** Histograms of the two groups
- *SPSS: Analyze → Nonparametric Tests → Independent Samples*

Paired t-test

- Dependent variable: Continuous (at least interval)
- Independent variable: Time point 1 or 2/ condition
- ***Use:***
  - A paired samples t-test can only be used when the data is paired or matched.
  - The test assesses whether the mean of the paired differences is zero.
- ***Plot:*** Histogram of differences
- *SPSS: Analyze → Compare means → Paired-samples T-test*

Assumptions	**How to check**	**What to do if assumption is not met**
Normality: paired differences* should be normally distributed	Histogram of differences / Shapiro Wilk	Wilcoxon signed rank

Wilcoxon Signed Rank test

- Non-parametric equivalent to the paired t-test
- Dependent variable: Ordinal/ Continuous
- Independent variable: Time/ Condition (binary)
- ***Use:***
  - The Wilcoxon signed rank test is used to compare two related samples, matched samples or repeated measurements on a single sample to assess whether their population mean ranks differ.
  - The absolute differences are ranked then the signs of the actual differences used to add the negative and positive ranks.
- ***Plot:*** Histogram of differences
- ***Interpretation:***
  - If p-value < 0.05 then there is evidence that the population mean ranks differ. Report the medians of the two sets of measurements
- *SPSS: Analyse → Nonparametric tests → Legacy dialogs → related samples*

One-way ANOVA

- Dependent variable: Continuous
- Independent variable: Categorical (at least 3 categories)
- ***Use:***
  - Used to detect the difference in means of 3 or more independent groups.
  - ANOVA uses the ratio of the between group variance to the within group variance to decide whether there are statistically significant differences between the groups or not.

Assumptions	**How to check**	**What to do if assumption is not met**
Residuals should be normally distributed	Histogram/ QQ plot of residuals / SW	Kruskall-Wallis test (non-parametric)
Homogeneity of variance	Levene's test / Bartlett's test	Welch test instead of ANOVA (adjusted for the differences in variance) and Games-Howell post hoc or Kruskall-Wallis

Kruskal-Wallis test

- Non-parametric equivalent to the one-way ANOVA
- Dependent variable: Ordinal/ Continuous
- Independent variable: Categorical
- ***Use:***
  - Kruskal-Wallis compares the medians of two or more samples to determine if the samples have come from different populations.
  - It is an extension of the Mann–Whitney U test to 3 or more groups.
  - The distributions do not have to be normal and the variances do not have to be equal.
- *SPSS: Analyze → Nonparametric tests → Independent samples*

Two-way ANOVA

- Dependent variable: Continuous
- Independent variables: Two categorical (2+ levels within each)
- ***Use:***
  - Comparing means for combinations of two independent categorical variables (factors).
- *SPSS: Analyse → General Linear Model → Univariate*

Assumptions	**How to check**	**What to do if assumption is not met**
Residuals should be normally distributed	Use the 'Save' menu within GLM to request the residuals and then use 'Explore' to produce histogram/ QQ plot of residuals / Shapiro Wilk test	Transform the data
Homogeneity of variance	Levene's test / Bartlett's test	Compare p-values with a smaller significance level e.g. 0.01

Chi-squared test

- Dependent variable: Categorical
- Independent variable: Categorical
- ***Use:***
  - The null hypothesis is that there is no relationship/association between the two categorical variables.
  - The chi-squared test compares expected frequencies, assuming the null is true, with the observed frequencies from the study.
  - When obtaining a significant chi-squared result, calculate percentages in a table to summarise where the differences between the groups are.
- ***Plot:*** Stacked/ multiple bar chart with percentages
- Assumptions:
  - 80% of expected cell counts >5
  - No cells with expected frequency below 1
- *SPSS: Analyze → Descriptive → Crosstabs → Statistics*

Correlation

- ***Use:***
  - Correlation (r) is used to measure the strength of association between two variables and ranges between -1 (perfect negative correlation) to 1 (perfect positive correlation).
- ***Plot:*** Scatterplot
- *SPSS: Analyse → Correlate → Bivariate Correlation*

Correlation coefficient value			Association
-0.3 to +0.3			Weak
-0.5 to -0.3	or	0.3 to 0.5	Moderate
-0.9 to -0.5	or	0.5 to 0.9	Strong
-1.0 to -0.9	or	0.9 to 1.0	Very strong

Cohen, L. (1992). Power Primer. Psychological Bulletin, 112(1) 155-159

Pearson’s correlation coefficient

- Dependent variable: Continuous
- Independent variable: Continuous
- Pearson’s correlation coefficient is the most common measure of correlation.
  - rho (ρ) = population correlation and r = sample correlation

Assumptions	**How to check**	**What to do if assumption is not met**
Continuous data for each variable	Check data	If ordinal data use Spearman’s or Kendall tau
Linearly related variables	Scatter plot	Transform data
Both variables are normally distributed	Histograms of variables/ Shapiro Wilk	Use rank correlation: Spearman’s or Kendall tau

Ranked correlation coefficients

- Dependent variable: Continuous/ Ordinal
- Independent variable: Continuous/ Ordinal
- **Spearman's Rank Correlation Coefficient**
  - Spearman's rank correlation coefficient is a non-parametric statistical measure of the strength of a monotonic relationship between paired data.
  - The notation used for the sample correlation is r_s

Assumptions	How to check	What to do if assumption is not met
Linearly related variables	Scatter plot	Transform data

Ranked correlation coefficients

- Dependent variable: Continuous/ Ordinal
- Independent variable: Continuous/ Ordinal
- **Kendall’s Tau Rank Corrélation Coefficient**
  - Use for small data sets with a large number of tied ranks
- ***Use:***
  - Kendall’s tau rank correlation coefficient is used to measure the association between two measured quantities.
  - A tau test is a non-parametric hypothesis test for statistical dependence based on the tau coefficient. Specifically, it is a measure of rank correlation, i.e. the similarity of the orderings of the data when ranked by each of the quantities.

Reliability

- Reliability can be divided into two main sections.
  - Reliability between raters (interrater agreement)
    - The technique for assessing agreement between raters/ instruments depends on the type of variable being compared.
      - For nominal data, **Cohen's kappa**
      - For scale data, the **Intraclass Correlation Coefficient** (ICC)
    - Data should be entered with one column for each rater and one row for each subject being rated.
    - Intrarater reliability is when measurements from the same person are being compared.
  - Reliability of a set of questions when measuring an underlying variable (Cronbach's alpha).

Cohen's Kappa

- ***Use:***
  - Assessing agreement between raters for categorical variables.
  - Kappa compares the proportion of actual agreement between raters (P_A) to the proportion expected to agree by chance (P_C).
  - The P_C values use expected values calculated as in the Chi-squared test for association (row total x column total/grand total).

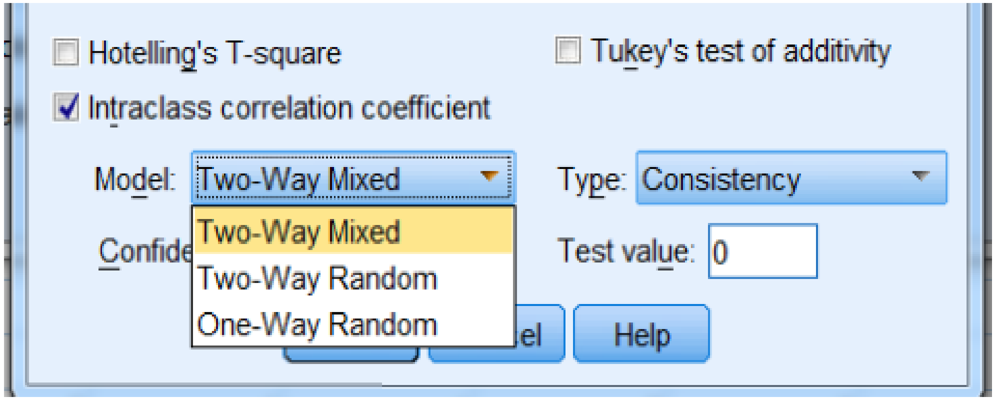
$$\text{kappa} = \frac{\mathrm{P_A} - P_C}{1 - P_C}$$

- ***Interpretation:***
  - There is a test for agreement which tests the hypothesis that agreement is 0 but like correlation, the interpretation of the coefficient itself is more important. The following guidelines were devised by Landis and Koch (1977).
- *SPSS: Analyse → Descriptive Statistics → Crosstabs*
  - Select 'Kappa' from the statistics options

Cohen's kappa	Strength of agreement
< 0	Poor (Agreement worse than by chance)
0 - 0.2	Slight
0.21 - 0.4	Fair
0.41 – 0.6	Moderate
0.61 – 0.8	Good
0.81 - 1	Very good

Intraclass Correlation Coefficient

- ***Use:***
  - A measure of agreement of continuous measurements for two or more raters. There are several options for the ICC in SPSS.
  - To choose the right combination of model, type and form, you need to ask the following questions:
    - Do all the raters rate all the subjects?
    - Are the raters the only possible raters (whole population) or a sample of raters?
    - Do raters need to match exactly or just be consistent (so raters may rank the subjects the same even if their scores don't match)?
    - Would you normally use just one rater or take an average of several raters?
- *SPSS: Analyse → Scale → Reliability Analysis*
  - In the 'Statistics' options, select 'Intraclass correlation coefficient'



Models:

One way random	Not all raters have scored all the subjects
Two way random	Random selection of raters and subjects (default)
Two way mixed	Analysis contains whole population of raters

Type:

Absolute agreement	Exact matches on scores/ measurements are required
Consistency	Raters are consistent with their scoring e.g. rater A consistently scores lower than rater B

Interpretation:

Interpret in the same way as Cohen's kappa.

Cronbach's alpha (reliability of scales)

- Researchers often use sets of likert style questions to measure an underlying latent variable that cannot be measured exactly.
- The set of questions is called a scale and the individual questions are called items.
- The scores for the items can be added or averaged to give an overall score for the scale. It is important that the items are all measuring the same underlying variable.
- ***Use:***
  - Cronbach’s alpha is a measure of internal consistency (how closely related the items are as a group). If the questions relate to the same issue, participants will be expected to get similar scores on each question.
- This measure is not robust against missing data.

- ***Interpretation:***
  - Cronbach's alpha ranges from 0 to 1 and scores are expected to be between 0.7 and 0.9.
  - Below is a commonly accepted rule of thumb for interpreting Cronbach's alpha.
- *SPSS: Analyse → Scale → Reliability Analysis.*

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Very high consistency (the items are so similar that some may not be needed)
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Acceptable
$\alpha < 0.7$	Poor internal consistency

Show video

Choosing a Statistical test

A very useful reference for advanced stuff….

Guide for selecting statistical techniques for analyzing social science data, 2nd ed

https://www.psc.isr.umich.edu/dis/infoserv/isrpub/pdf/GuideSelectingStatisticalTechniques_OCR.PDF

http://www.dataanalytics.org.uk/Data%20Analysis/Statistics/choosing-your-stats-test.htm

https://onlinecourses.science.psu.edu/stat500/node/67/

Available Statistical Packages

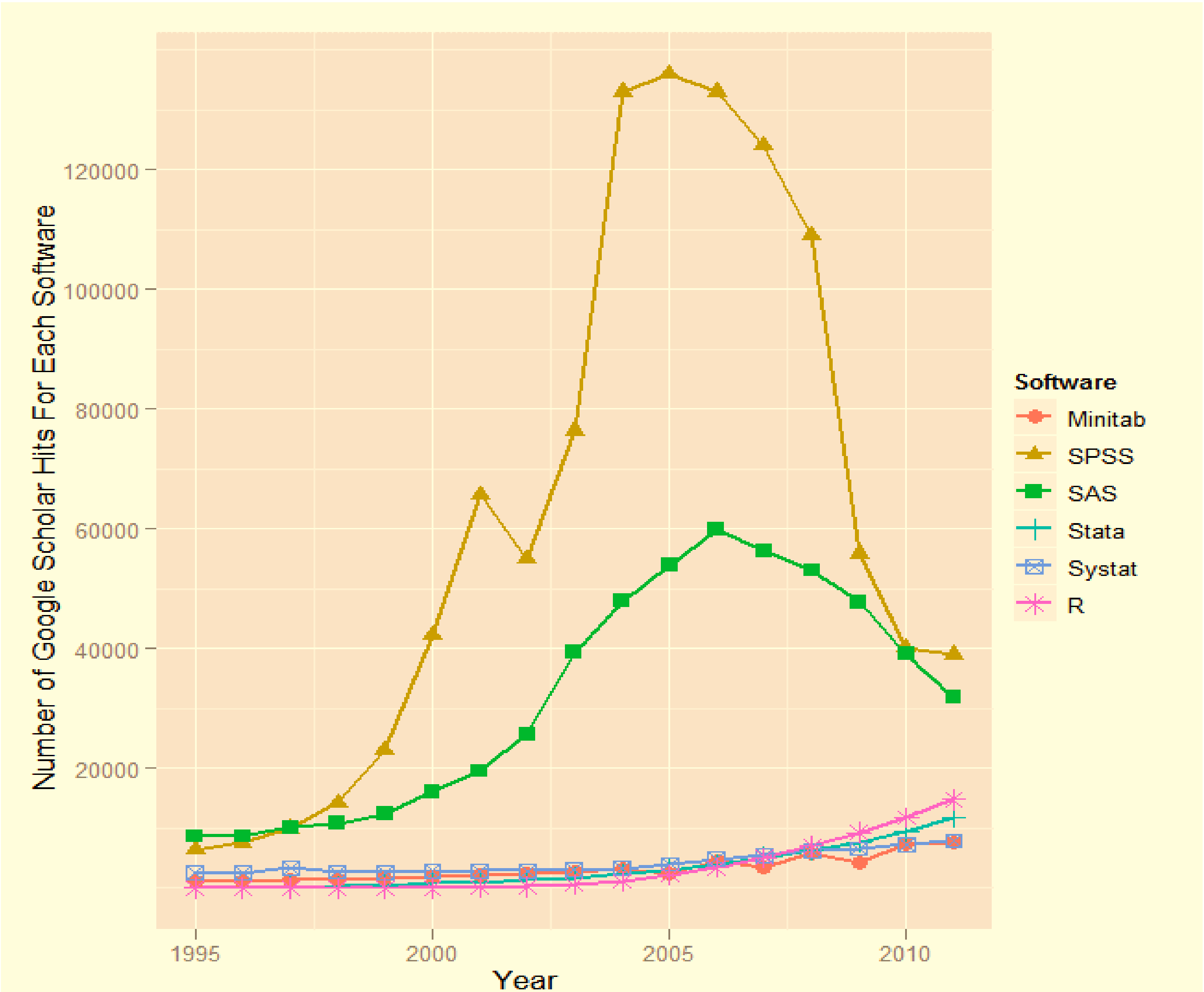
Commercial

- Excel
- SPSS
- MINITAB
- SAS

Free Software

- LibreOffice Calc
- PSPP
- R

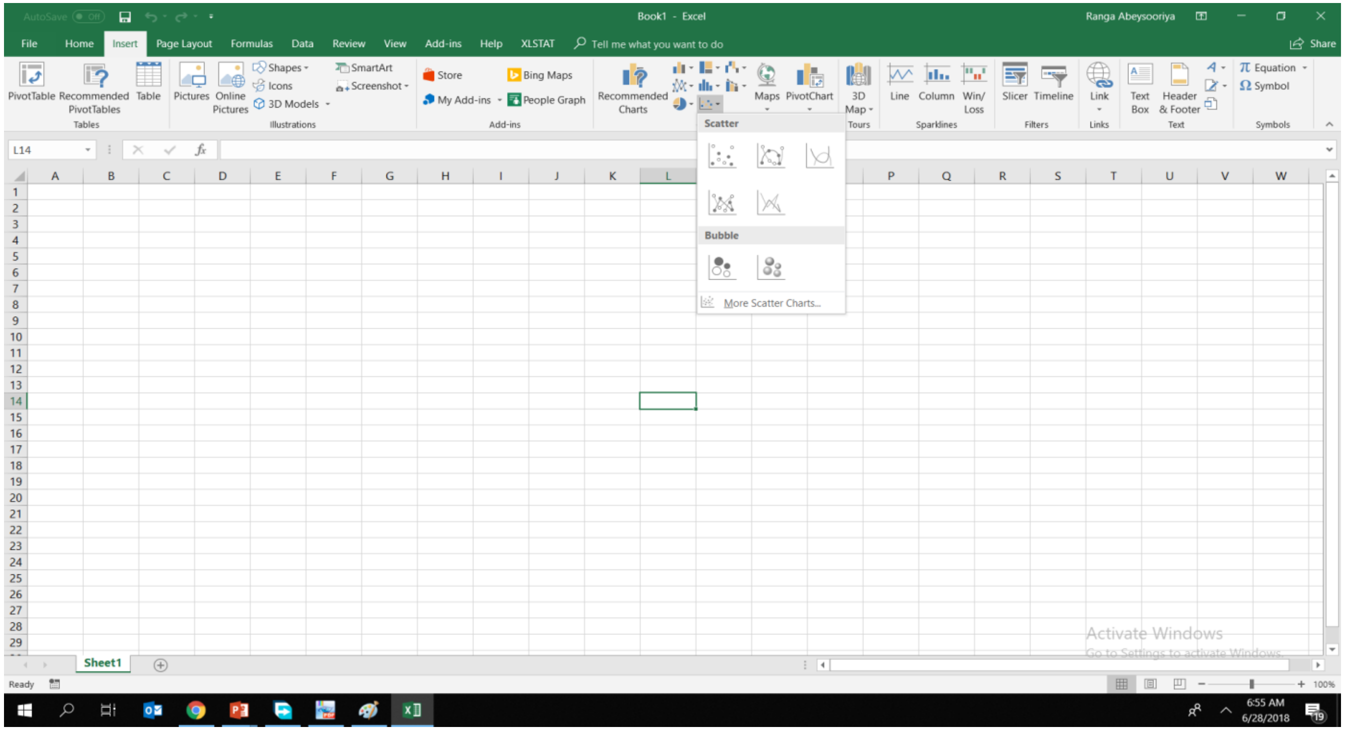
Usage of each packages within (Academia)

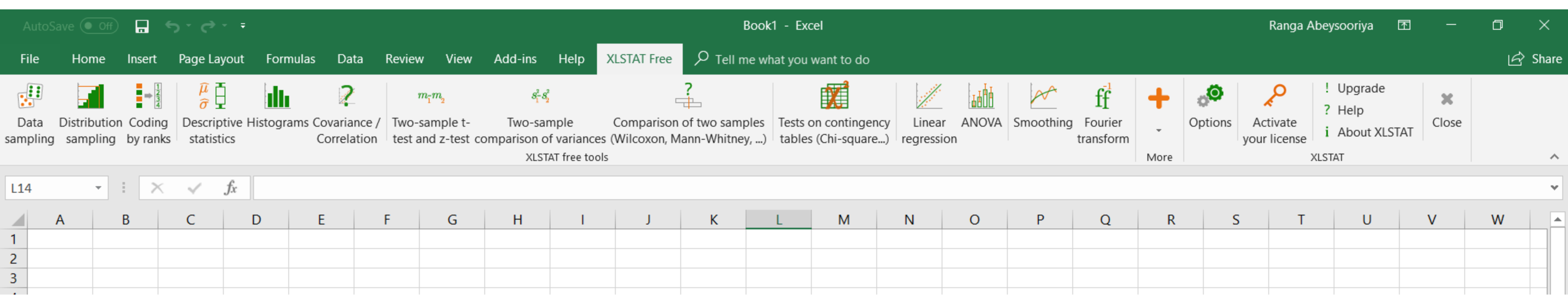


Use of data analysis software in academic publications as measured by hits on Google Scholar.

http://www.pitt.edu/~super7/47011-48001/47651.ppt

Microsoft Excel





PROS

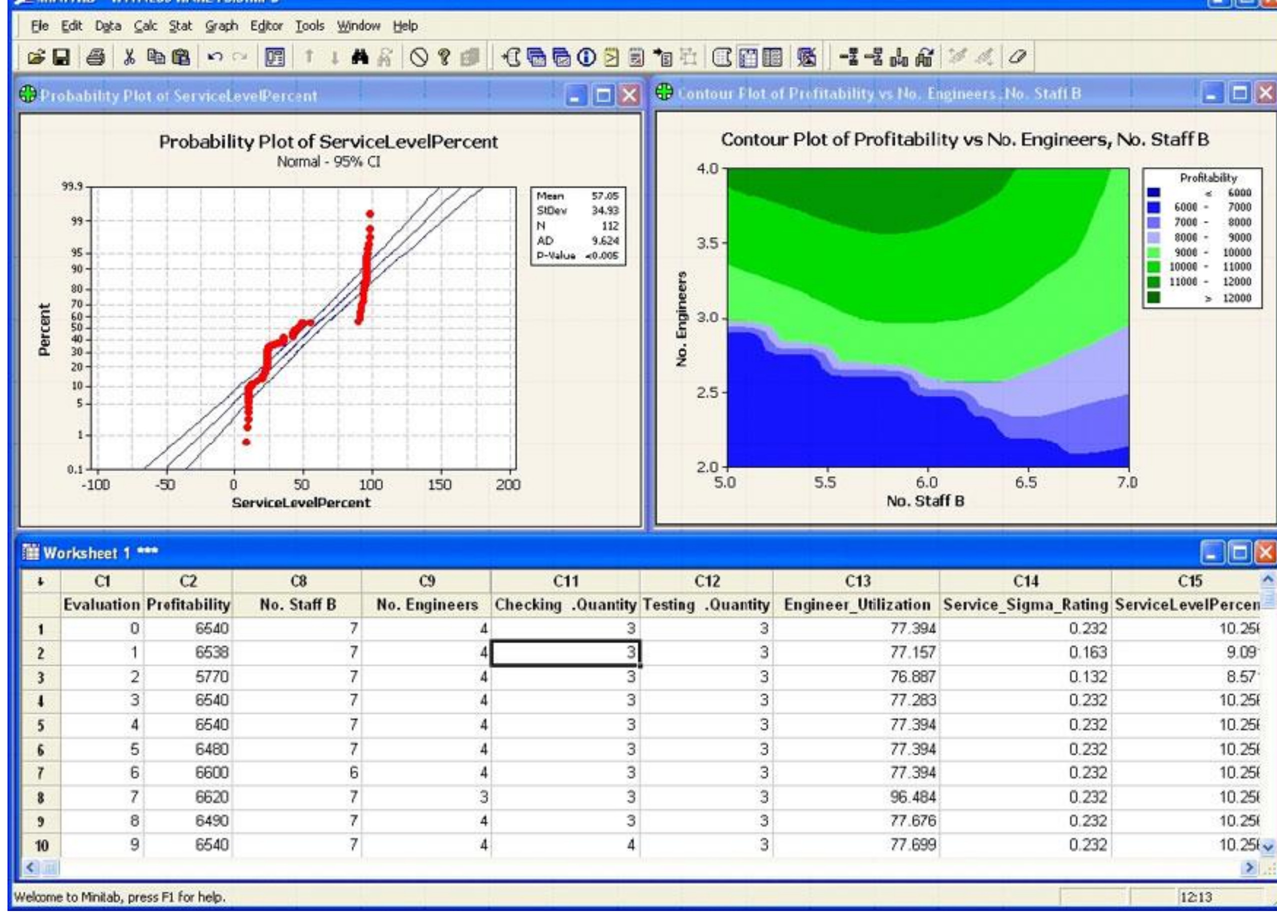
- Nearly ubiquitous and is often pre-installed on new computers
- User friendly
- Very good for basic descriptive statistics, charts and plots
- Add-ins

CONS

- Costs money
- Not powerful for advanced stuff

http://www.pitt.edu/~super7/47011-48001/47651.ppt

Minitab



PROS

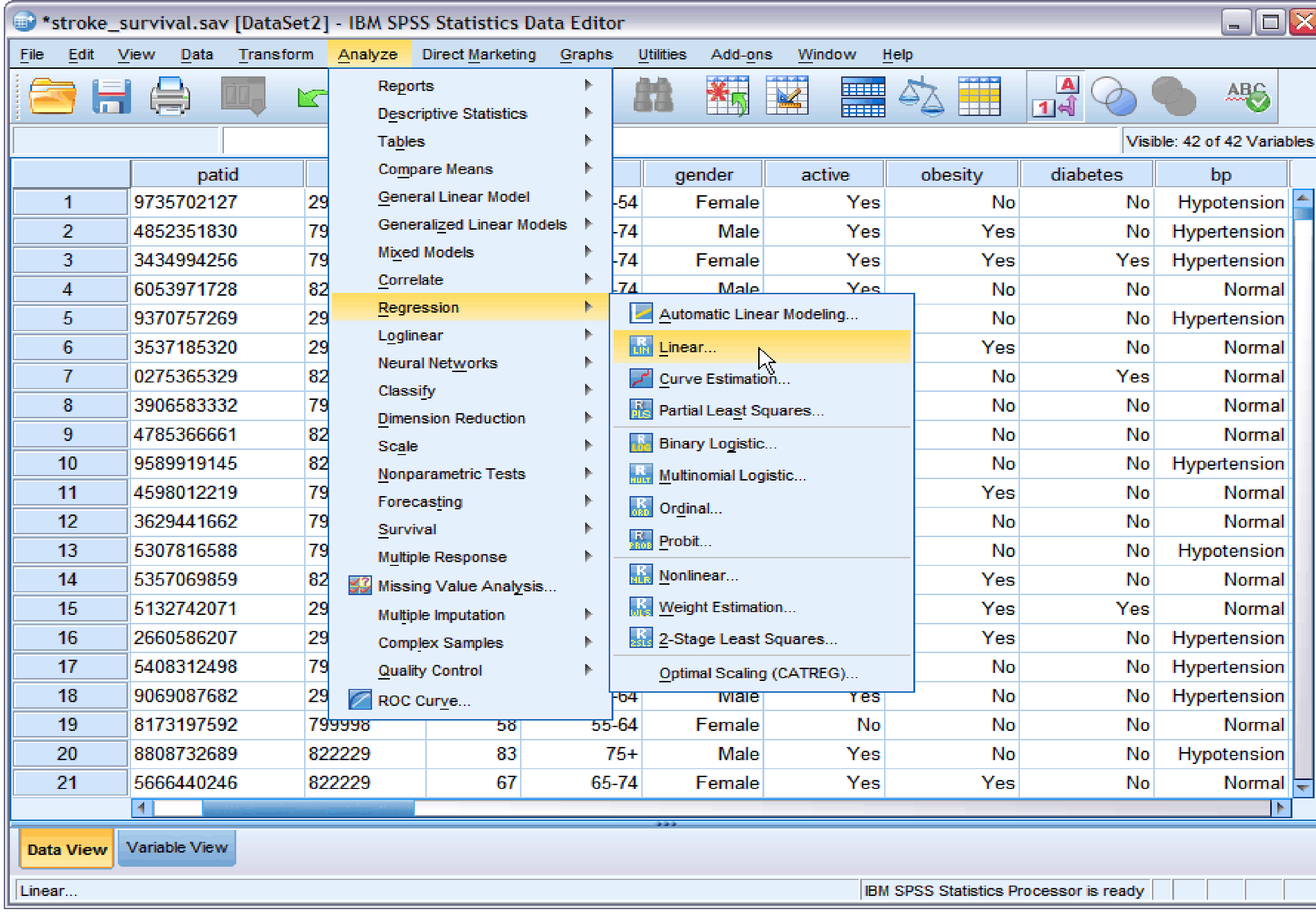
- Easy to learn and use
- Often taught in schools in introductory statistics courses
- Widely used in engineering for process improvement

CONS

- Costs Money
- Not suitable for very complicated statistical computation and analysis
- Not often used in academic research

http://www.pitt.edu/~super7/47011-48001/47651.ppt

SPSS



PROS

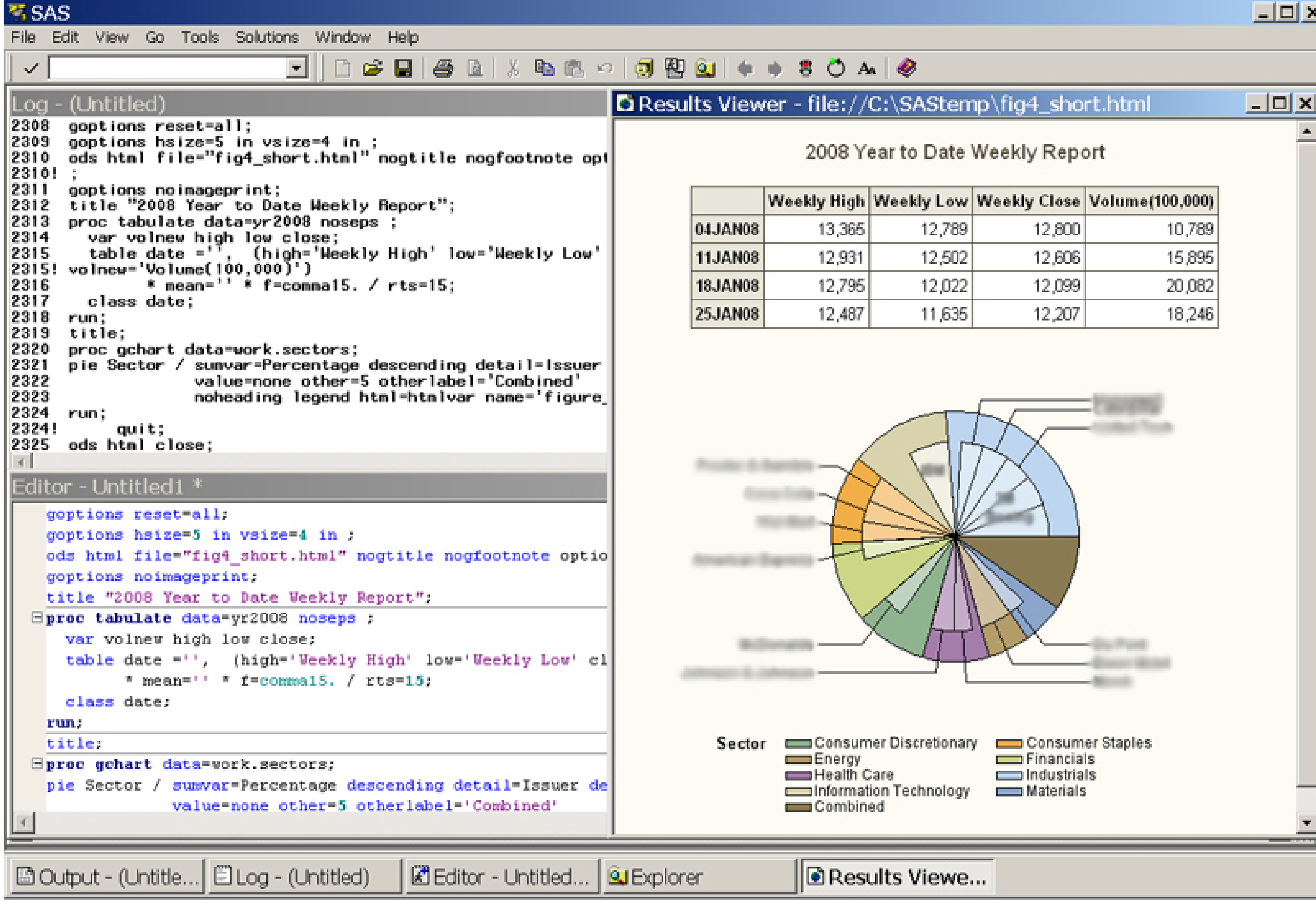
- Easy to learn and use
- More powerful than Minitab
- One of the most widely used statistical packages in academia and industry
- Has a command line interface in addition to menu driven user interface
- One of the most powerful statistical package that is also easy to use.

CONS

- Very expensive
- Not adequate for modeling and cutting edge statistical analysis

http://www.pitt.edu/~super7/47011-48001/47651.ppt

SAS



PROS

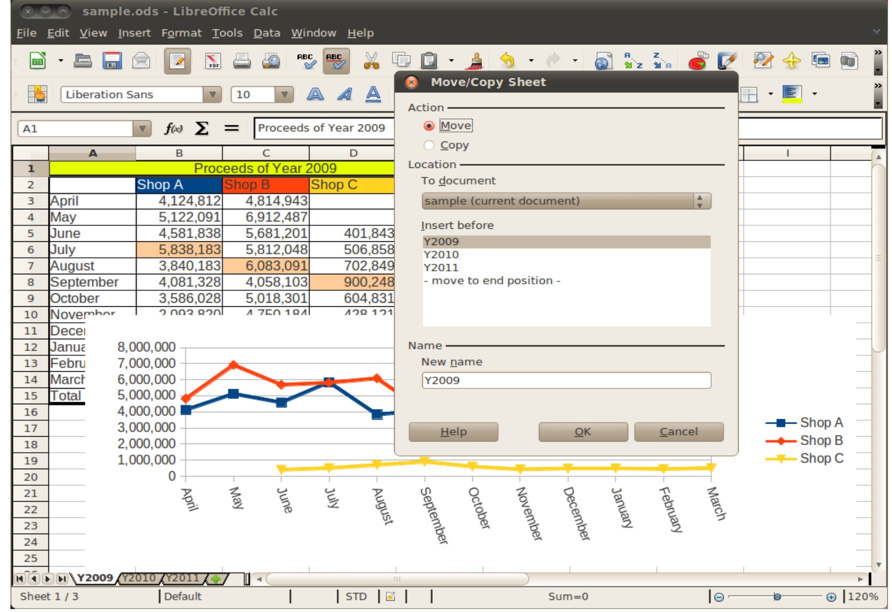
- Widely accepted as the leader in statistical analysis and modeling
- Widely used in the industry and academia
- Very flexible and very powerful.

CONS

- Very expensive
- Not user friendly
- Steep learning curve

http://www.pitt.edu/~super7/47011-48001/47651.ppt

LibreOffice Calc



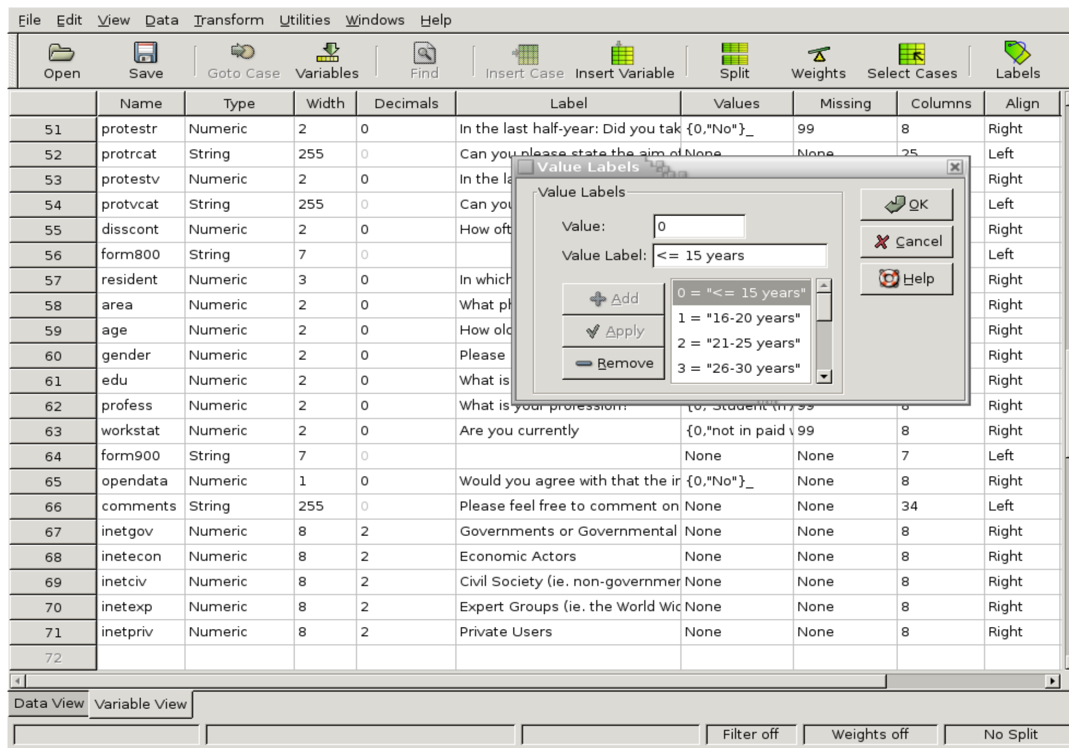
PSPP

COST

- Free

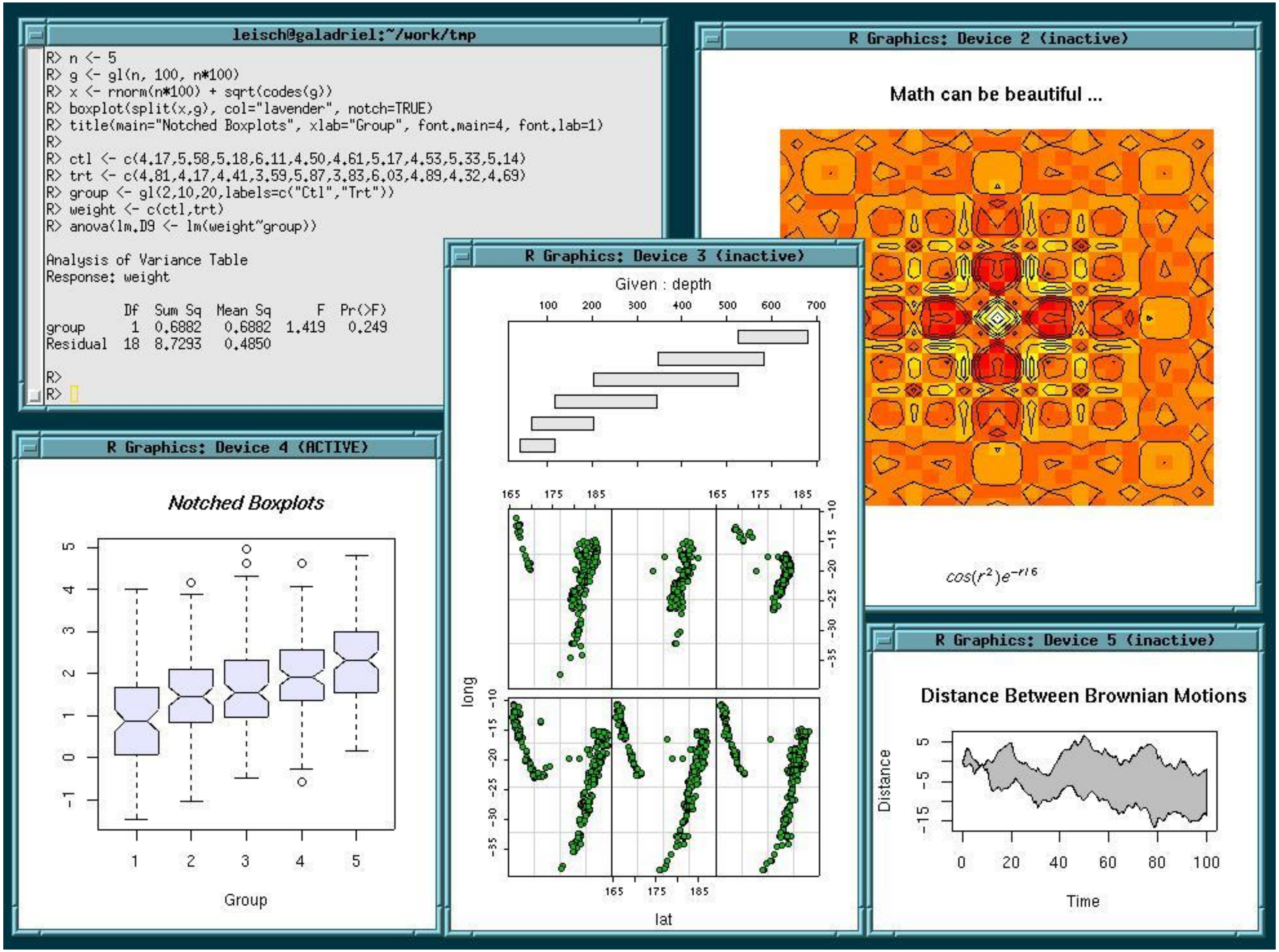
PRO

- Aims as a free SPSS alternative with an interface that closely resembles SPSS
- User friendly
- Good enough for basic statistical analysis



http://www.pitt.edu/~super7/47011-48001/47651.ppt

R


leisch@galadriel:~/work/tmp
R> n <- 5
R> g <- gl(n, 100, n*100)
R> x <- rnorm(n*100) + sqrt(codes(g))
R> boxplot(split(x,g), col="lavender", notch=TRUE)
R> title(main="Notched Boxplots", xlab="Group", font.main=4, font.lab=1)
R>
R> ctl <- c(4.17,5.58,5.18,6.11,4.50,4.61,5.17,4.53,5.33,5.14)
R> trt <- c(4.81,4.17,4.41,3.59,5.87,3.83,6.03,4.89,4.32,4.69)
R> group <- gl(2,10,20,labels=c("Ctl","Trt"))
R> weight <- c(ctl,trt)
R> anova(lm.D9 <- lm(weight~group))
Analysis of Variance Table
Response: weight
Df Sum Sq Mean Sq F Pr(>F)
group 1 0.6882 0.6882 1.419 0.249
Residual 18 8.7293 0.4850
R>
R>
R Graphics: Device 2 (inactive)
Math can be beautiful ...
$cos(r^2)e^{-r/6}$
R Graphics: Device 3 (inactive)
Given : depth
100 200 300 400 500 600 700
165 175 185
long
lat
R Graphics: Device 4 (ACTIVE)
Notched Boxplots
Group
R Graphics: Device 5 (inactive)
Distance Between Brownian Motions
Distance
Time

Research Methods

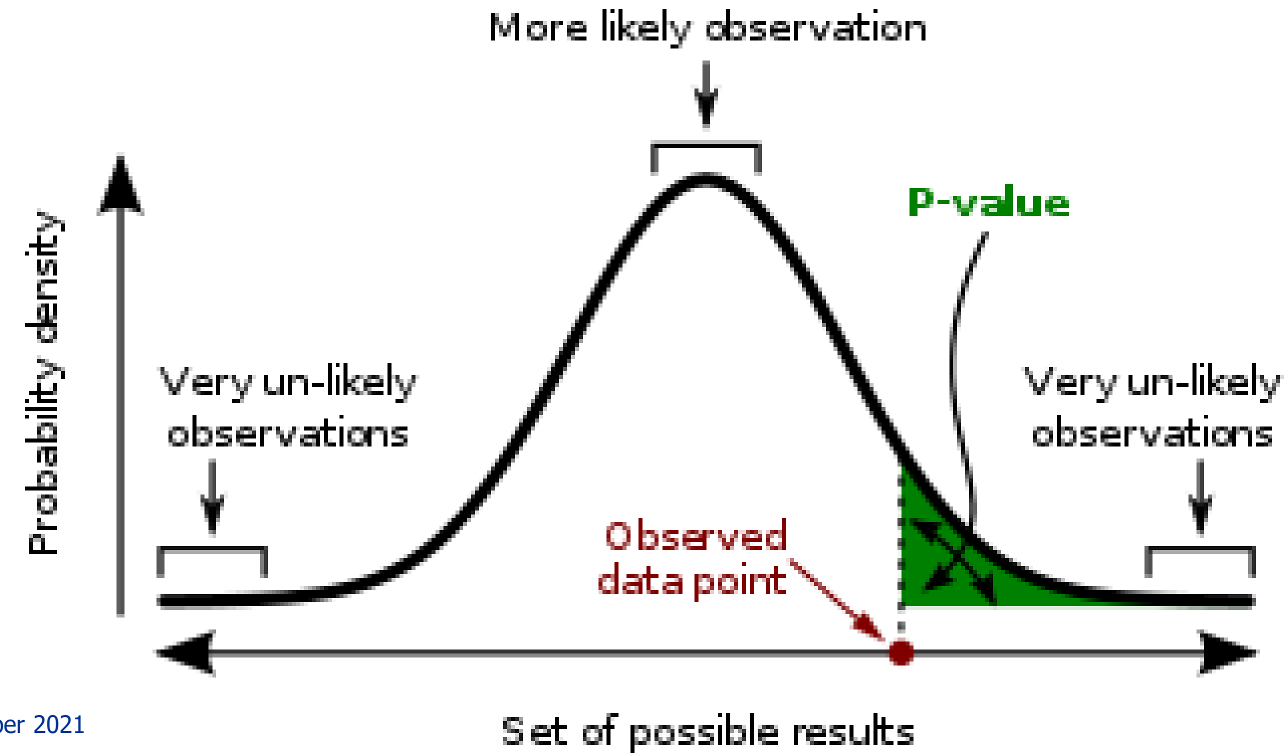
Lecture 5-2: Results interpretation

Agenda

- P-value
- Types of errors
- Interpretation of box plots
- Common hypothesis tests – Mean, proportions and variances
- Contingency table example: Chi square test
- Comparison of means of two independent samples -t test example
- Interpretation of SPSS output

P-value

- **p-value or probability value is the probability that if null hypothesis were true, sampling variation would produce an estimate that is further away from the hypothesized value than the data estimate**



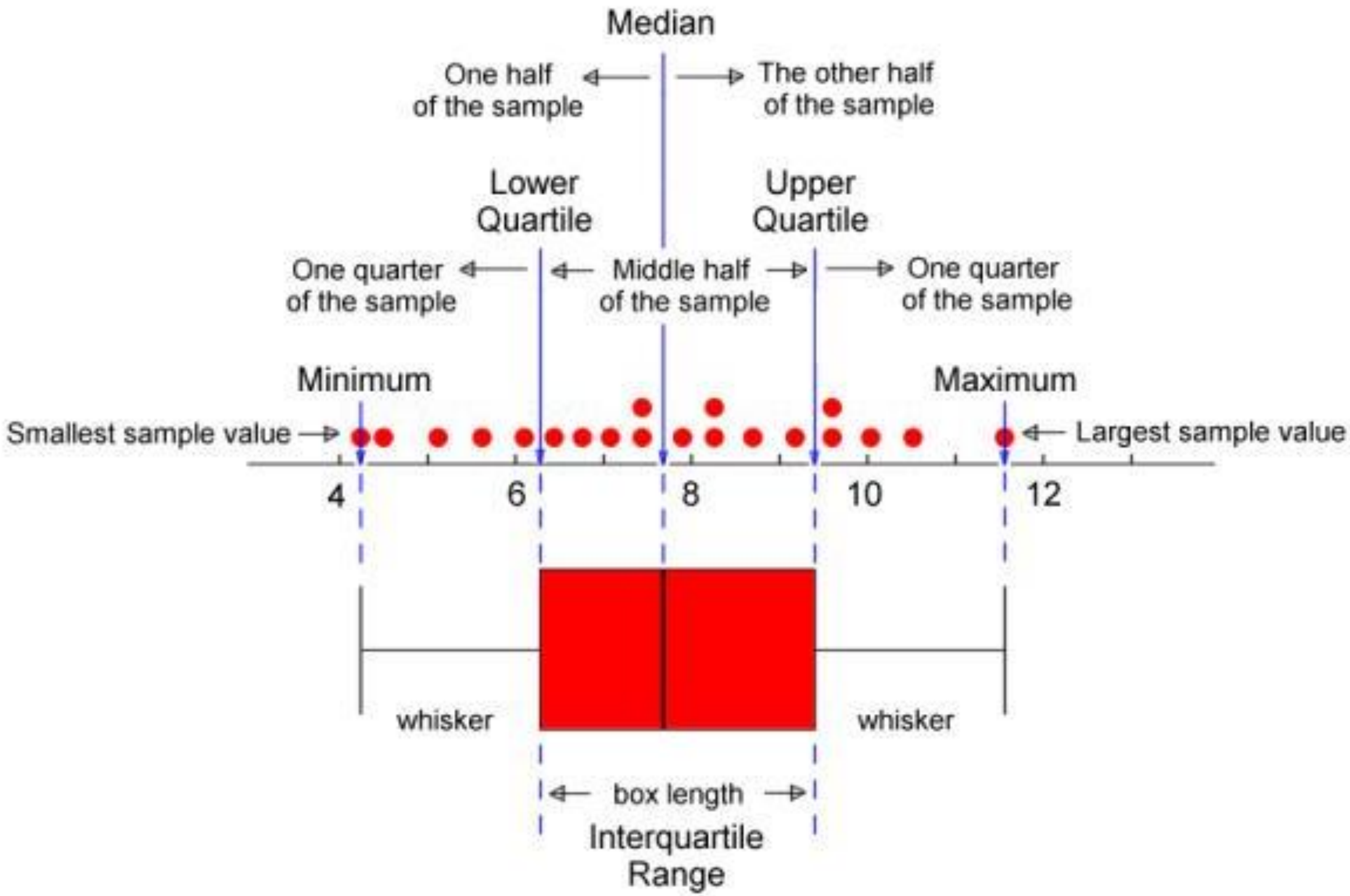
Types of errors

- **There are two types of inferential errors we might make:**
  - –Type I : rejecting the null hypothesis when it is in fact true; a false positive
  - –Type II : not rejecting the null hypothesis when it is in fact false; a false negative

	Null hypothesis H_0 is really . . .	
Action taken	True	False
Reject	Type I error committed	success
Don't reject	success	Type II error committed

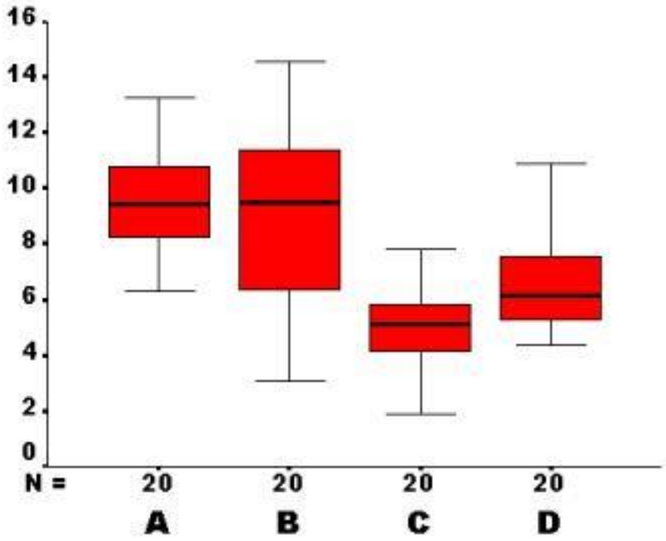
Interpretation of box plots

- **Inter-quartile range**
  The middle “box” represents the middle 50% of scores for the group. The range of scores from lower to upper quartile is referred to as the inter-quartile range. The middle 50% of scores fall within the inter-quartile range.
- **Upper quartile**
  Seventy-five percent of the scores fall below the upper quartile.
- **Lower quartile**
  Twenty-five percent of scores fall below the lower quartile.
- **Whiskers**
  The upper and lower whiskers represent scores outside the middle 50%. Whiskers often (but not always) stretch over a wider range of scores than the middle quartile groups.


Median
One half of the sample
The other half of the sample
Lower Quartile
Upper Quartile
One quarter of the sample
Middle half of the sample
One quarter of the sample
Minimum
Maximum
Smallest sample value
Largest sample value
4
6
8
10
12
whisker
whisker
box length
Interquartile Range

Example of box plot interpretation

- In the above plot, sample A and B appear to have similar centres, which exceed those of C and D.
- Sample B appears to have larger variability than the other three samples.
- Samples A, B and C are reasonably symmetric, but sample D is skewed to the right.
- There are no obvious outliers in any of the samples.



SLIIT
COMPUTING | BUSINESS | ENGINEERING

Common hypothesis tests-Mean

No	Type of test	Nature of means	Variance of population	Population/sample size	Hypothesis	Test used for acceptance
1	One tailed	Single mean	known	Population size is infinite	$H_0: \mu < k$; $H_1: \mu > k$ or $H_0: \mu > k$; $H_1: \mu < k$	$Z_{\bar{X}} = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$, $Z_{\bar{X}} < Z_{\alpha}$ or $Z_{\bar{X}} > -Z_{\alpha}$
2	Two tailed	Single mean	known	Population size is infinite	$H_0: \mu = k$; $H_1: \mu \neq k$	$Z_{\bar{X}} = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ $-Z_{\alpha/2} < Z_{\bar{X}} < Z_{\alpha/2}$
3	One tailed	Single mean	known	Population size is finite	$H_0: \mu < k$; $H_1: \mu > k$ or $H_0: \mu > k$; $H_1: \mu < k$	$Z_{\bar{X}} = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}}$ $Z_{\bar{X}} < Z_{\alpha}$ or $Z_{\bar{X}} > -Z_{\alpha}$
4	Two tailed	Single mean	known	Population size is finite	$H_0: \mu = k$; $H_1: \mu \neq k$	$Z_{\bar{X}} = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}}$ $-Z_{\alpha/2} < Z_{\bar{X}} < Z_{\alpha/2}$
5	One tailed	Single mean	unknown	Sample size is large	$H_0: \mu < k$; $H_1: \mu > k$ or $H_0: \mu > k$; $H_1: \mu < k$	$Z_{\bar{X}} = \dfrac{\bar{X} - \mu}{S / \sqrt{n}}$ *inf inite population* $Z_{\bar{X}} = \dfrac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}}$ *finite population* $Z_{\bar{X}} < Z_{\alpha}$ or $Z_{\bar{X}} > -Z_{\alpha}$

Common hypothesis tests-Mean

No	Type of test	Nature of means	Variance of population	Population/sample size	Hypothesis	Test used for acceptance
6	Two tailed	Single mean	unknown	Sample size is large	$H_0: \mu = k$; $H_1: \mu \neq k$	$Z_{\bar{X}} = \frac{\bar{X} - \mu}{S / \sqrt{n}}$ *infinite population* $Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}$ *finite population* $-Z_{\alpha/2} < Z_{\bar{X}} < Z_{\alpha/2}$
7	One tailed	Single mean	unknown	Sample size is small (<30)	$H_0: \mu < k$; $H_1: \mu > k$ or $H_0: \mu > k$; $H_1: \mu < k$	$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$ n-1 degrees of freedom if n<30 $t > -t_{\alpha}$ *or* $t < t_{\alpha}$
8	Two tailed	Single mean	unknown	Sample size is small (<30)	$H_0: \mu = k$; $H_1: \mu \neq k$	$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$ n-1 degrees of freedom if n<30 $-t_{\alpha/2} < t < t_{\alpha/2}$
9	One tailed	Difference between 2 means	known	----	$H_0: \mu_1 < \mu_2$; $H_1: \mu_1 > \mu_2$ or $H_0: \mu_1 > \mu_2$; $H_1: \mu_1 < \mu_2$	$Z_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $Z_{\bar{X}} < Z_{\alpha}$ *or* $Z_{\bar{X}} > -Z_{\alpha}$
10	Two tailed	Difference between 2 means	known	----	$H_0: \mu_1 = \mu_2$; $H_1: \mu_1 \neq \mu_2$	$Z_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $-Z_{\alpha/2} < Z_{\bar{X}} < Z_{\alpha/2}$

SLIIT COMPUTING | BUSINESS | ENGINEERING

Common hypothesis tests-Mean

No	Type of test	Nature of means	Variance of population	Population/sample size	Hypothesis	Test used for acceptance
11	One tailed	Difference between 2 means	unknown	Sample sizes are large	$H_0: \mu_1 < \mu_2$; $H_1: \mu_1 > \mu_2$ or $H_0: \mu_1 > \mu_2$; $H_1: \mu_1 < \mu_2$	$Z_{\bar{X}_1-\bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $Z_{\bar{X}} < Z_\alpha \quad or \quad Z_{\bar{X}} > -Z_\alpha$
12	Two tailed	Difference between 2 means	unknown	Sample sizes are large	$H_0: \mu_1 = \mu_2$; $H_1: \mu_1 \neq \mu_2$	$Z_{\bar{X}_1-\bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $-Z_{\alpha/2} < Z_{\bar{X}} < Z_{\alpha/2}$
13	One tailed	Difference between 2 means	unknown	Sample sizes are small (<30)	$H_0: \mu_1 < \mu_2$; $H_1: \mu_1 > \mu_2$ or $H_0: \mu_1 > \mu_2$; $H_1: \mu_1 < \mu_2$	$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n2}}}$ assume that $\sigma_1 = \sigma_2$ $t > -t_\alpha \quad or \quad t < t_\alpha$
14	Two tailed	Difference between 2 means	unknown	Sample sizes are small (<30)	$H_0: \mu_1 = \mu_2$; $H_1: \mu_1 \neq \mu_2$	$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n2}}}$ assume that $\sigma_1 = \sigma_2$ $-t_{\alpha/2} < t < t_{\alpha/2}$

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Common hypothesis tests-proportions

No.	Type of test	Type of proportion	Hypothesis	Test used for acceptance
1	One tailed	Single proportion	$H_0: p < k; \quad H_1: p > k$ or $H_0: p > k; \quad H_1: p < k$	$Z_{\bar{p}} = \dfrac{\bar{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ $Z_{\bar{p}} < Z_{\alpha} \quad or \quad Z_{\bar{p}} > -Z_{\alpha}$ For large sample size (n>30), binomial distribution can be approximated to normal distribution
2	Two tailed	Single proportion	$H_0: p = k; \quad H_1: p \neq k$	$Z_{\bar{p}} = \dfrac{\bar{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ $-Z_{\alpha/2} < Z_{\bar{p}} < Z_{\alpha/2}$ For large sample size (n>30), binomial distribution can be approximated to normal distribution
3	One tailed	Two proportions	$H_0: p_1 < p_2; \quad H_1: p_1 > p_2$ or $H_0: p_1 > p_2; \quad H_1: p_1 < p_2$	$Z_{\bar{p}_1 - \bar{p}_2} = \dfrac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$ $Z_{\bar{p}_1 - \bar{p}_2} < Z_{\alpha} \quad or \quad Z_{\bar{p}_1 - \bar{p}_2} > -Z_{\alpha}$
4	Two tailed	Two proportions	$H_0: p_1 = p_2; \quad H_1: p_1 \neq p_2$	$Z_{\bar{p}_1 - \bar{p}_2} = \dfrac{\bar{p}_1 - \bar{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$ $-Z_{\alpha/2} < Z_{\bar{p}_1 - \bar{p}_2} < Z_{\alpha/2}$

Common hypothesis tests-Variances

No.	Type of test	Type of population	Hypothesis	Test used for acceptance
1	One tailed	Single population variance	$H_0: \sigma < k$; $H_1: \sigma > k$ or $H_0: \sigma > k$; $H_1: \sigma < k$	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ $\chi^2 < \chi^2_{n-1}(1-\alpha)$ *or* $\chi^2 < \chi^2_{n-1}(\alpha)$
2	Two tailed	Single population variance	$H_0: \sigma = k$; $H_1: \sigma \neq k$	$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ $\chi^2_{n-1}(1-\alpha/2) < \chi^2 < \chi^2_{n-1}(\alpha/2)$
3	One tailed	Two population variances	$H_0: \sigma_1 < \sigma_2$; $H_1: \sigma_1 > \sigma p_2$ or $H_0: \sigma_1 > \sigma_2$; $H_1: \sigma_1 < \sigma_2$	$F = \frac{S_1^2}{S_2^2}$ assume that $\sigma_1 = \sigma_2$ $F < F^{n_1-1}_{n_2-1}(\alpha)$ *or* $F^{n_1-1}_{n_2-1}(1-\alpha) < F$
4	Two tailed	Two population variances	$H_0: \sigma_1 = \sigma_2$; $H_1: \sigma_1 \neq \sigma_2$	$F = \frac{S_1^2}{S_2^2}$ assume that $\sigma_1 = \sigma_2$ $F^{n_1-1}_{n_2-1}(1-\alpha/2) < F < F^{n_1-1}_{n_2-1}(\alpha/2)$

Nominal variables

- It is required to investigate the effect of education level in text adoption. The observed frequencies are given in the table below

	Academic programme type		
	BA	BA+MA	BA+MA+PhD
Adopt		25	35
Do not adopt	23	24	1
Differ	5	5	10

- Contingency table:
  - Independent variable: education level - Nominal
  - Dependent Variables: text adoption - Nominal
- Hypothesis
  - Null hypothesis H_o: There is no effect of education level on text adoption against
  - The alternative H_1: education level has an effect on text adoption

Contingency table

		Academic Program Type			
		BA only	*BA+MA*	*MA, MA, and PhD*	
Adoption Decision	*Adopt*	18 (24.58)	25 (28.85)	35 (24.58)	78
	Do not adopt	23 (15.12)	24 (17.75)	1 (15.12)	48
	Defer	5 (6.30)	5 (7.40)	10 (6.30)	20
		46	54	46	*N* = 146

$$\text{Chi}^2 = \frac{-6.58^2}{24.57} + \frac{-3.85^2}{28.85} + \frac{10.42^2}{24.58} + \frac{7.88^2}{15.12} + \frac{6.25^2}{17.75} + \frac{-14.12^2}{15.12} + \frac{-1.30^2}{6.30} + \frac{-2.40^2}{7.40} + \frac{3.70^2}{6.30}$$

$$= 1.76 + .51 + 4.42 + 4.10 + 2.20 + 13.19 + .27 + .78 + 2.17 = 29.40$$

Determining the significance of Chi-square

- The degrees of freedom for a contingency table is equal to (columns - 1) * (rows - 1), or (3-1)*(3-1) = 4. The critical value of Chi^2 for *p* = .05 and 4 d.f. is 9.488.
- Values higher than this occur fewer than 5 times in 100 by chance alone. (95% confidence level)
- The observed value of Chi^2 is 31.32, sufficiently large to reject the null hypothesis.
- Consequently, we conclude that educational setting does affect instructors' decisions about text adoption.

Comparison of means of two independent samples -t test

- It is required to investigate the partner's stance on issues. Two groups have been selected such that one year or less from the marriage (group 1) and more than one year from the marriage (group 2). Then the perception score is determined for each subject in both groups and it is tabulated as follows.

Group 1

Score	f
6	3
7	4
8	8
9	10
10	5
11	5
12	6
13	4
14	4
15	4

Group 2

Score	f
8	2
9	1
10	2
11	5
12	7
13	8
14	13
15	12

- Independent Variable: group - Nominal
- Dependent Variable: score – interval (Interval/Ratio)
- Type of selection: Two Independent Groups

Comparison of means of two independent samples -t test

- Null hypothesis H_o: there is no difference between groups **against**
- The alternative H_1: group 2 has a better perception about the partner's stance on issues

Length of Marriage

One Year or Less				More than One Year			
Score	*f*	$(X-\bar{X}_1)^2$	$f\cdot(X-\bar{X}_1)^2$	Score	*f*	$(X-\bar{X}_1)^2$	$f\cdot(X-\bar{X}_1$
6	3	16	48				
7	4	9	36				
8	8	4	32	8	2	25	50
9	10	1	10	9	1	16	16
10	5	0	0	10	2	9	18
11	5	1	5	11	5	4	20
12	6	4	24	12	7	1	7
13	4	9	36	13	8	0	0
14	4	16	64	14	13	1	13
15	4	25	25	15	12	4	48

One Year or Less:

$\sum f\cdot X = 500$

$\bar{X}_1 = 500/50 = 10.00$

$\sum_{i=1}^{N} f\cdot(X_i-\bar{X}_1)^2 =$ 280.0

More than One Year:

$\sum f\cdot X = 650$

$\bar{X}_1 = 650/50 = 13.00$

$\sum_{j=1}^{N} f\cdot(X_i-\bar{X}_1)^2 =$ 172.0

$$t = \frac{(10-13)-0}{\sqrt{\dfrac{280+172}{50+50-2}\cdot\left(\dfrac{1}{50}+\dfrac{1}{50}\right)}}$$

$$= \frac{-3.00}{.430} = -6.98$$

Determining the significance of *t* for independent samples

- The degrees of freedom associated with this *t*-test are determined by $(N_1 + N_2) - 2$, or $50 + 50 - 2 = 98$.
- The closest *t* critical from the table of *t*-values for these d.f. and $p = .05$, 1-tail (for a directional hypothesis), is 1.671.
- Since the *t* observed is much greater in magnitude and in the correct (negative) direction, we can conclude that the longer people have been married, the more accurately they perceive their partner's stance on issues.

SLIIT
COMPUTING | BUSINESS | ENGINEERING

One-Sample z-Test.

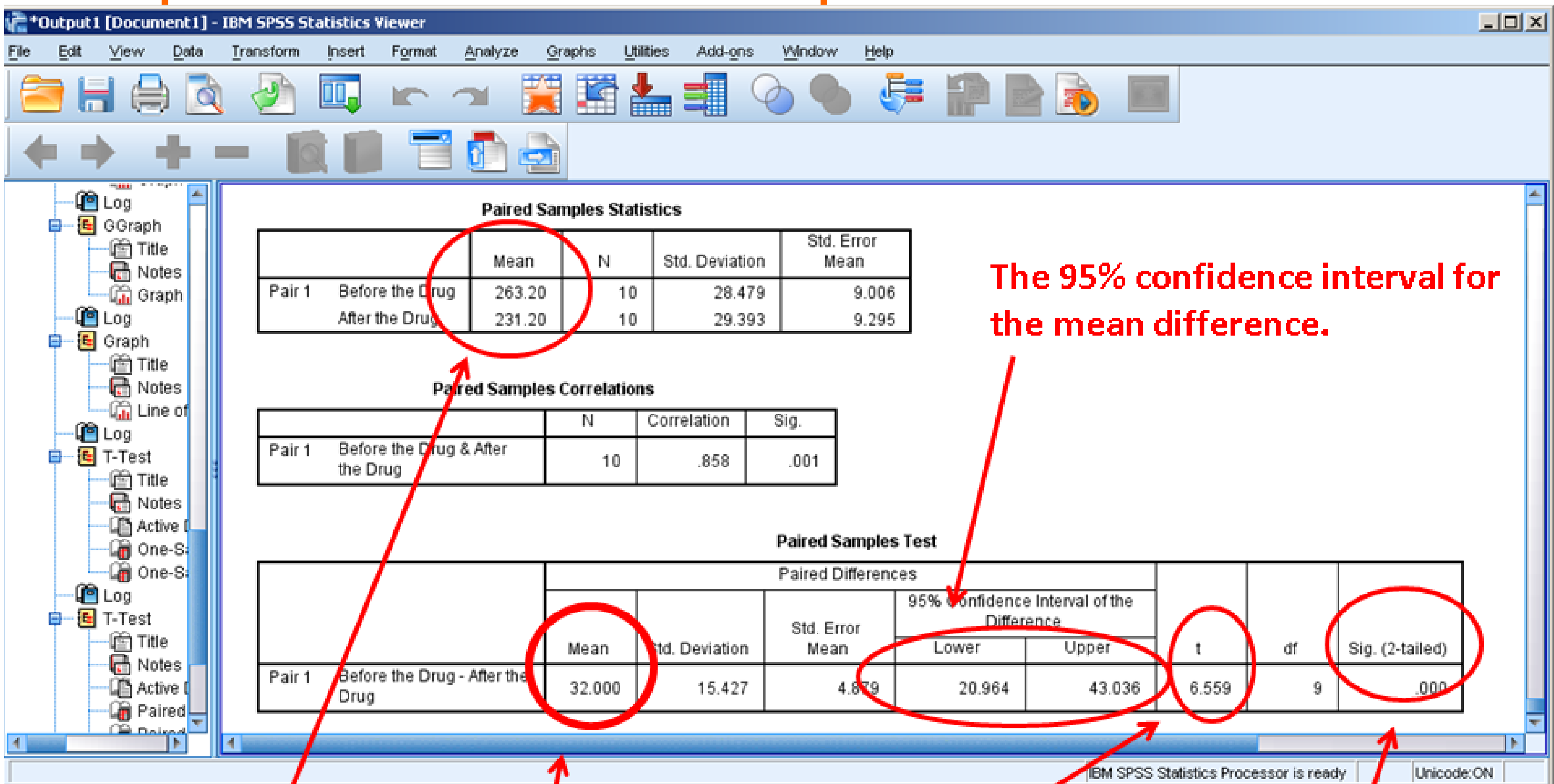
- A **one-sample z-test** is used to **test** whether a population parameter is significantly different from some hypothesized value.
- Example: In the population, the average IQ is 100 with a standard deviation of 15. A team of scientists wants to test a new medication to see if it has either a positive or negative effect on intelligence, or no effect at all. A sample of 30 participants who have taken the medication has a mean of 140. Did the medication affect intelligence, using alpha = 0.05?
- Null hypothesis H_o: $\mu = 100$ against the alternative H_1: $\mu \neq 100$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad Z = \frac{140 - 100}{\frac{15}{\sqrt{30}}} = \frac{40}{2.74} = 14.60$$

- Reject null hypothesis and conclude that Medication significantly affected intelligence, $z = 14.60$, $p < 0.05$

SLIIT
COMPUTING | BUSINESS | ENGINEERING

Interpretation of SPSS output-1



Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before the Drug	263.20	10	28.479	9.006
	After the Drug	231.20	10	29.393	9.295

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Before the Drug & After the Drug	10	.858	.001

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference: Lower	Upper			
Pair 1	Before the Drug - After the Drug	32.000	15.427	4.879	20.964	43.036	6.559	9	.000

The 95% confidence interval for the mean difference.

The Sample Means

The Test Statistic

The Mean Difference

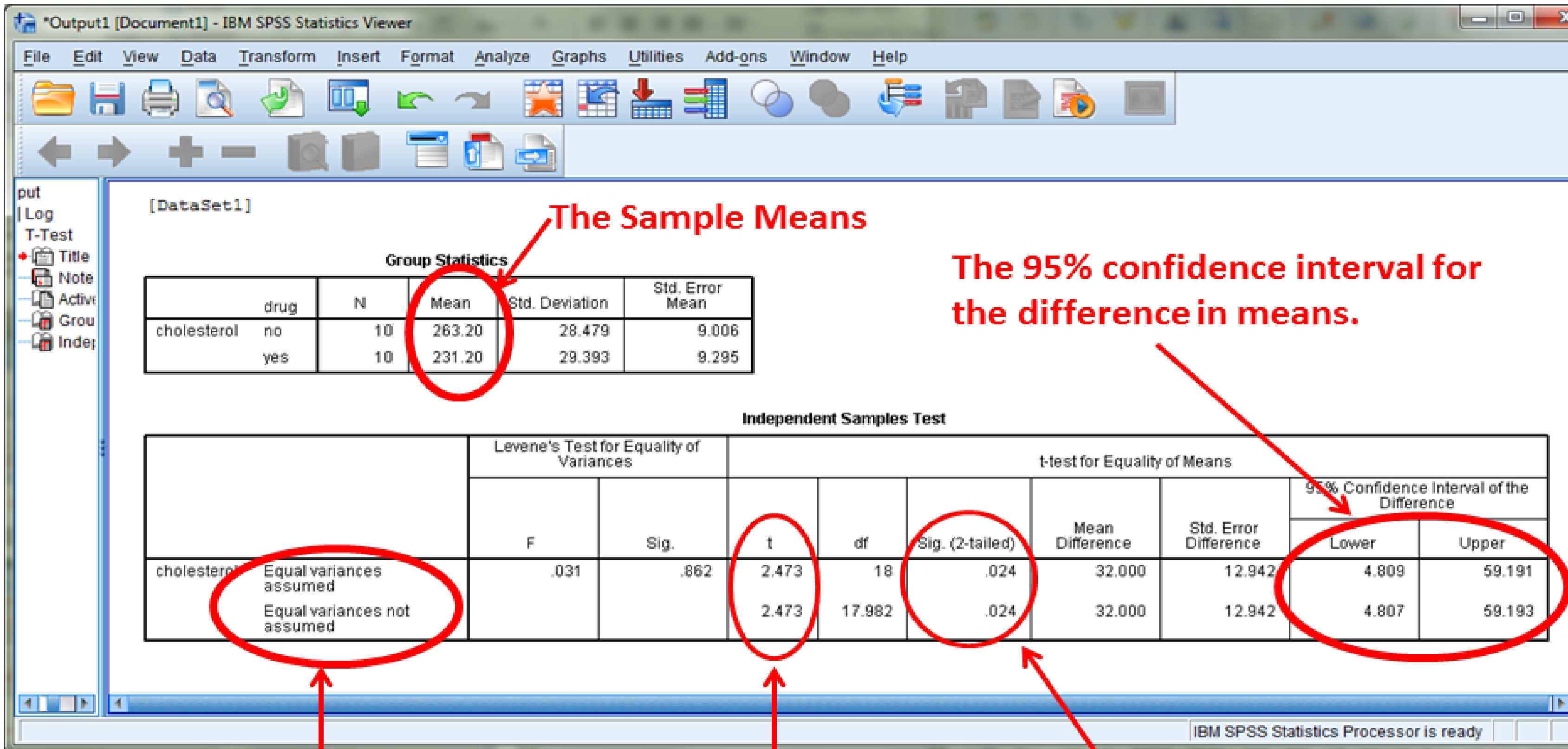
P-value (Two-Tailed)

Interpretation-1

- **The two tailed p-value is given as 0.000 (it is actually 0.0001)**
- **If one tailed test is done, get half of this p-value**
- **Regardless of one tailed test or two tailed test, reject the null hypothesis that mean difference is zero**
- **In confidence interval approach, at 95% confidence level mean difference is between 21 and 43.**

- **This is a paired test where there may be the same subjects prior to and after treatment or there is a relationship exists between two groups among subjects. Sample size should be equal for a paired test.**

Interpretation of SPSS output-2



Group Statistics

	drug	N	Mean	Std. Deviation	Std. Error Mean
cholesterol	no	10	263.20	28.479	9.006
	yes	10	231.20	29.393	9.295

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
cholesterol	Equal variances assumed	.031	.862	2.473	18	.024	32.000	12.942	4.809	59.191
	Equal variances not assumed			2.473	17.982	.024	32.000	12.942	4.807	59.193

Assumption on the Variance

The Test Statistic

P-value (Two-Tailed)

Interpretation-2

- **The two tailed p-value is given as 0.024**
- **Depending on the significant level, do not reject the null hypothesis of equal means**
- **Or one tail test, p-value would be 0.024/2=0.012.**
- **In confidence interval approach, it is 95% confident that the difference in population means is between 5 and 59. Big interval**
- **It does not matter whether we assume that variances are equal or not**

- **This is a Hypothesis Test to compare two Means and two samples contain Independent Data. The sample sizes may be different.**

Thank You..!