DA using MS Excel

(Research Methodology

Nabaraj Poudyal

Nabaraj Poudyal

M Sc Statistics

Assistant Professor

Tribhuvan University

Amrit Campus

What we will discuss

30 minutes

What Scientific
Research is

30 minutes

Types of Data

**Research
Methodology**

Data Analysis Using MS Excel

30 minutes
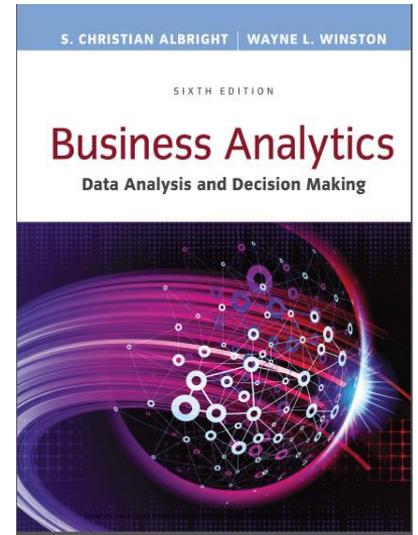
Why Data is needed/
Statistics

30 minutes

Concept of Data analysis

RM & DA

With today's technology, companies are able to collect tremendous amounts of data with relative ease. Indeed, many organization now have more data than they can handle. However, the data are usually meaningless until they are analyzed for trends, patterns, relationships, and other useful information. This training elaborate you a variety of methods, from simple to complex, to help you analyze data sets and uncover important information.

http://bijaylalpradhan.com.np/

S. CHRISTIAN ALBRIGHT | WAYNE L. WINSTON

SIXTH EDITION

Business Analytics
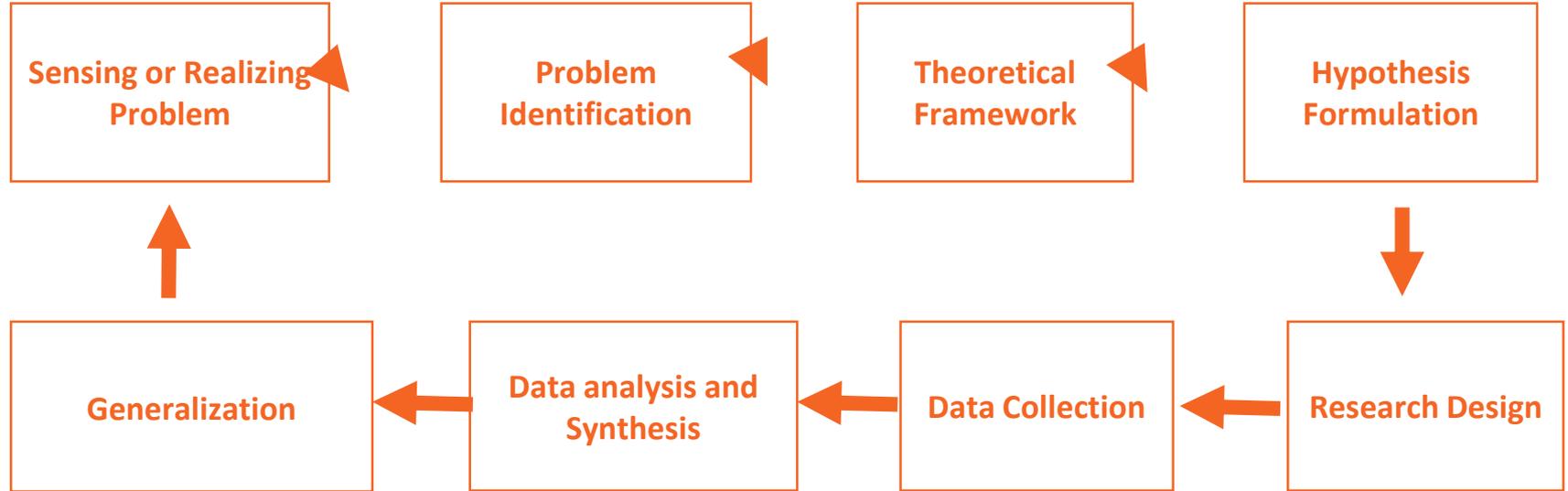
Data Analysis and Decision Making

What is Research???

- Research isn't information gathering:
  - Gathering information from resources such as books or magazines isn't research.
  - No contribution to new knowledge.

- Research isn't the transportation of facts:
  - Merely transporting facts from one resource to another doesn't constitute research.
  - No contribution to new knowledge although this might make existing knowledge more accessible.

Research is…

"…the systematic process of collecting and analyzing information (data) in order to increase our understanding of the phenomenon about which we are concerned or interested."

Process of Research

| Sensing or Realizing Problem | ← | Problem Identification | ← | Theoretical Framework | ← | Hypothesis Formulation |

| Generalization | ← | Data analysis and Synthesis | ← | Data Collection | ← | Research Design |

The Scientific Research Process

Research Problem

- Pre-occurred idea that generate necessity of research
- Idea generated through
  - discussion with experts
  - Going through literature
  - Involve in activities related to subject matter
- Generally express in negative statement
- Group Participation

Guidelines for selecting Research Problem

- Should be researcher's interest
- Should be chain of thinking
- It may be of verification of problem
- Should be in manageable limit
- Should have some utilities

Criteria of good problem

- Should have relationship among variables

- Should state clearly

- Should be such as to imply possibilities of empirical testing

Example: Objective

The broad objective of the study is to find the relationship among the awareness, perception and readiness of students for online learning.

However the study has the following specific objectives

- To assess the awareness level for online learning among Students

- To ascertain the attitude, perception, perceived usefulness and ease of use of online learning system

- To assess if the awareness, perception and readiness dimensions vary across gender, education, experience and age group.

Example: Objective

The broad objective of the study is to assess the problem of microfinance in promoting micro-enterprises for poverty alleviation of women.

However the study has the following specific objectives

- To enumerate the female beneficiaries of micro-enterprises in the study site
- To identify the patterns of investment of loans and returns to investment
- To analyse the procedure of obtaining loan from microfinance organization
- To make comparative study of income distribution among the beneficiaries and non beneficiaries of microfinance
- To study the loan repayment pattern

Hypothesis

- Two words
  - Hypo: Under
  - Thesis: reasoned theory
- Theory which is not fully reasoned
- Tentative answer of the research question
- Imaginate idea or guess depending upon previous accumulated knowledge which can be put to test to determine its validity.
- Generally specify relationship between variables or with specific value

Statistics

(1) Descriptive Statistics: In descriptive statistics, it deals with collection of data, its presentation in various forms, such as tables, graphs and diagrams and findings averages and other measures which would describe the data.

For Example: Industrial statistics, population statistics, trade statistics etc… Such as businessman make to use descriptive statistics in presenting their annual reports, final accounts, bank statements.

Statistics

(2) Inferential Statistics: In inferential statistics, it deals with techniques used for analysis of data, making the estimates and drawing conclusions from limited information taken on sample basis and testing the reliability of the estimates.

For Example: Suppose we want to have an idea about the percentage of illiterates in our country. We take a sample from the population and find the proportion of illiterates in the sample. This sample proportion with the help of probability enables us to make some inferences about the population proportion. This study belongs to inferential statistics

Descriptive and Inferential Statistics
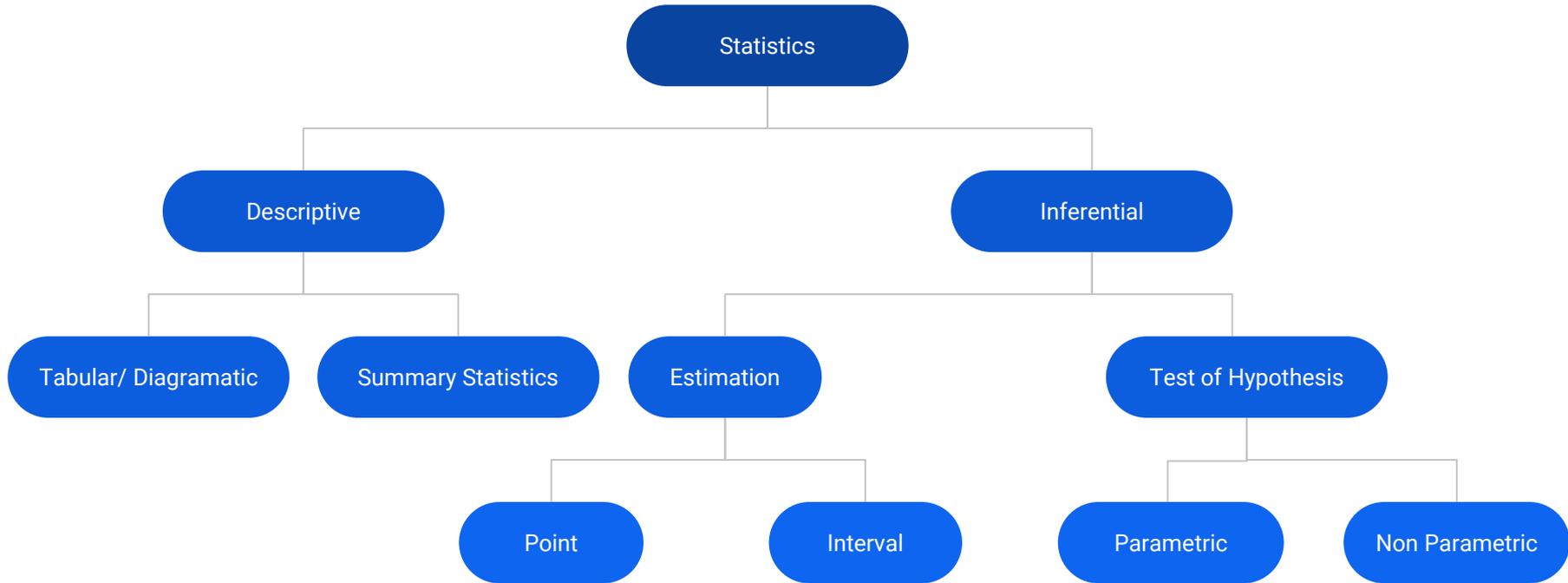
Uncertainty always remains while generalizing results from a sample to a population. The degree of uncertainty is measured in terms of *probability* in inferential statistics.

Data And Its Type

○ A characteristic or measurement that may different from one entity to another or place to place or time to time is called Data, which is able to distinguish among them. For eg. The measurement for height, weight, income, expenditure, demand etc.

○ Data are collected for an investigation or research depending on the nature of the problem, they may relate to individuals, families, houses, village, business etc. The collected data are known as observations.

○ Observations may be measure of the characteristics or the numerical facts

Data and its Type

- **Categorial Data: C**ategorical data are defined by some quality characteristic such as gender, eye colour, etc. Categorical data is further classified as either nominal or ordinal. Nominal data defines some group characteristic such as gender, nationality, profession, membership of clubs or societies, etc. Ordinal data is the result of ranking products, goods, services; opinions, etc. in order of preference. For example, viewers may be asked to rank television programmes in order of preference:1,2,3,etc. But the difference between ranks 1 and 2 may not be the same as between ranks 2 and 3.

Data and its Type

Numeric data obtained **n**umerically by counts or measurements. Numeric data may be discrete or continuous. Discrete data can only assume certain distinct values; for example the numbers that turns when a die is thrown can only be 1, 2, 3, 4, 5 or 6. Continuous data on the other hand can assume any value from a continuous set of values: for example the time between arrivals at a check-in desk could be a value such as 3.25 minutes.

Numerical data may be further classified as ratio or interval.

Data and its Type

An interval scale is based on measurements from an arbitrary zero – a classic example being temperature scale in Fahrenheit with zero at 32∘F where, unlike ranks, a difference of say, 10∘F between any two temperatures is the same whether it is between 40∘F and 50∘F or between 85∘F and 95∘F. A ratio scale is an interval scale with a natural zero so that measurements can be compared. Ratio scales include distances, heights, weights volumes, time. For example, a queuing time of 40 minutes is twice as long as a queuing time of 20 minutes.

Data and its type

- The process of arranging **data** into homogenous groups or classes according to some common characteristics present in the **data** is called **classification**. In other word classification is a partition or a sub partition of total possible outcomes into different distinct groups or elements. For example: During the process of sorting letters in a post office, the letters are **classified** according to the cities and further arranged according to streets.

Data and its types

- The data, which are collected according to time variation (year, month, quarter, week, day, hours, minutes etc) are called **Time series data.**

- The data referring to a single time point or a single space point (or any single factor of the variable/attribute is a cross section data.)

- **Pooled data** is a mixture of time series data and cross-section data. One example is GNP per capita of all South Asian countries over ten years.

Sources of Primary and Secondary Data

- Primary
  - Data collected by investigator from personal experimental studies for a specific research
  - First hand data
  - Collected when secondary data are unavailable and inappropriate
- Secondary
  - The data (published or unpublished form which has collected by others for their purpose) can be utilized for study of another investigator, such data is said to be secondary data.

Sources of Primary and Secondary Data

- **Source of Primary data**

  - Questionnaire survey (post, internet)

  - Interview (personal/telephone)

  - Focus group discussion

  - Observation

  - Key informants interview

Sources of Primary and Secondary Data

- **Source of Secondary data**
  - Usual public sources
    - Nepal census of Household and Population, agriculture, business, vital statistics etc
    - Governmental organization-national and district level use for development of society (office of ministry, municipality, district development office etc)
    - Opinion and poll taken by others
    - Health and microbial survey done by others
   - INGO's, NGO's, UN publication
  - Unusual sources: Easily accessible
    - The yellow pages, Newspapers, Bulletin Board, Films, Post cards, old prints, Topographical maps etc

Problems on collecting data

- Problems in collecting primary data
  - Timeframe, budgetary
  - Transportation
  - Non response error
  - Biasness of enumerator
  - Lack of expertise in construction of questionnaire and collection of data
- Problems in collecting secondary data
  - Definition of terms and units
  - If two set data comparison may make confusion
  - Data may not be exact form of requirement
  - Reliability and suitability

Data Collection Techniques
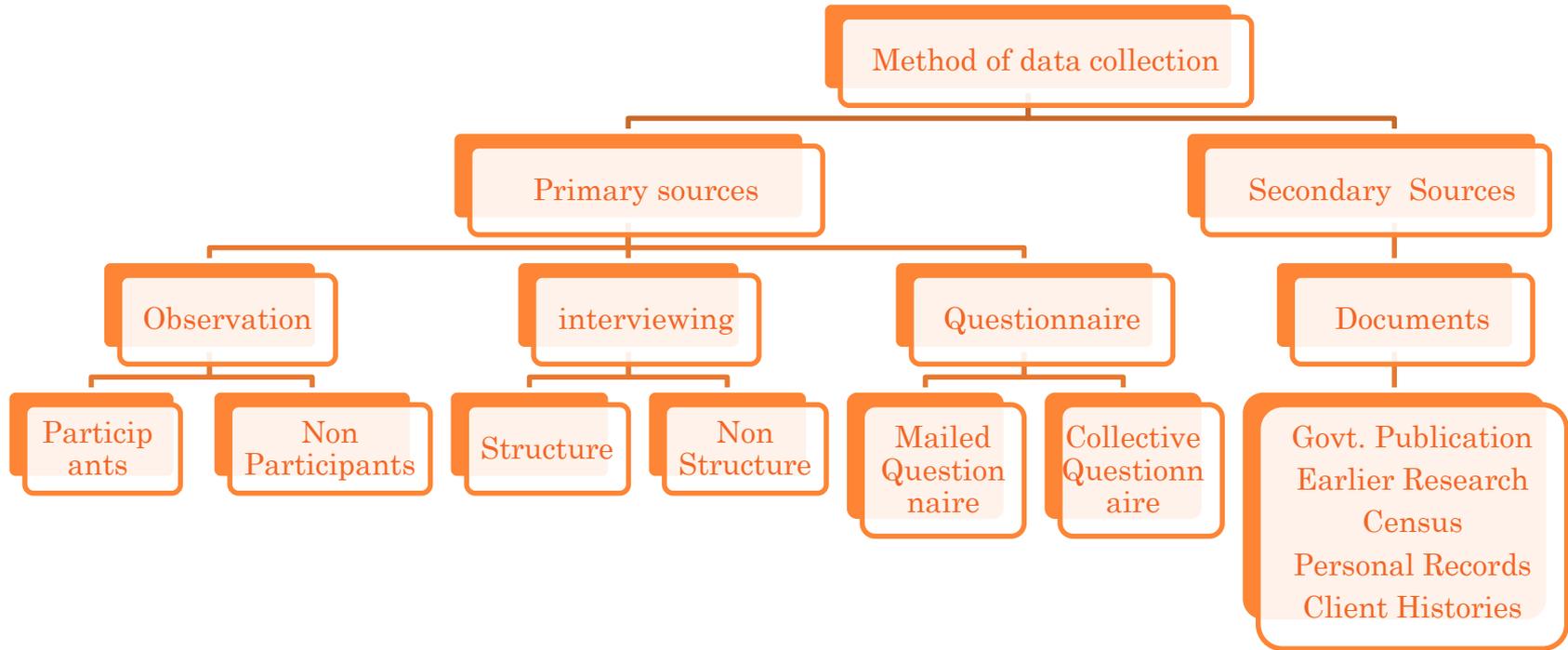
Creating excel spreadsheet

- Unique variable names should be in the first row
- Data should be in column format
- Data in the same column should be of the same data type
- Some data validation features are available
- Data audit features are available for existing spreadsheets

Categorical / Continuous Data

ID	age	gender	height	treatment	disease
1	20	0	70	1	0
2	33	1	66	0	0
3	56	1	62	1	1
4	45	1	65	0	1
5	29	0	72	1	0
6	52	0	68	1	1
7	41	1	68	0	1

Preparing files for statistical analysis

- Allow adequate time for data preparation
- Better the quality of data, less time analyzing
  - Know your data
    - Look at frequency distributions and scatterplots
  - Multiple checks for errors
  - Minimize missing data if at all possible
    - Be aware of amount of data missing and why
- Freeze the dataset
  - Copy to another file and date the file
  - Document any corrections made to file and also correct in original database and on forms
- Plan on recoding categorical variables so each group has a sufficient sample size
- Prepare a separate code sheet for data

Preparing files for statistical analysis

- General spreadsheet design
  - One line header row with a unique one word name for each variable
  - Do not mix data types within one column
  - Unique identifying number for each case
  - Only include raw, un-summarized data, i.e., no summary statistics or graphs in spreadsheet
  - Date format with four digit years
  - Avoid underlining, bold fonts, or italics
  - Do not leave blank rows or columns in between data
  - Do not use a row to label a group, use a grouping variable (column)

Preparing files for statistical analysis

- Missing Data
- Consider what software will be used for analysis
- Use different codes to indicate reason missing
- e.g., not applicable, unable to complete, or missing
- If numeric field
- Must not be a valid data point
- Do not use text, such as "NA", "missing", "*"

Summary

- Well designed systems minimize data errors and future problems
- Data management systems should be chosen based on resources and individual needs
  - Spreadsheets are appropriate for small and simple data sets
  - Databases provide more options for data management
- Add simple validations to check data entry
- Following guidelines for preparing files for statistical analysis will save time
- Data transfer software is available to transfer data between applications
- Limit PHI and keep data secure

What we can do using Technology

Data Collection

Designing Questionnaire

Online Survey: Google form

MS form, Survey Monkey,

KOBO tools

Data arrangement Analysis

Excel, SPSS, Stata, R, Python………………..

Presentation / Dissemination

Diagramatic form

Textual/ Tabular form

Report presentaiton

Good luck!

We hope you'll use these tips to go out and deliver a memorable pitch for your product or service!

For more (free) presentation tips relevant to other types of messages, go to heathbrothers.com/presentations