DATA COLLECTION AND ANALYSIS

Types of data

By Nature: Quantitative or Qualitative

By timeframe:

- **Cross Section Data- Data values observed at a fixed point in time**
- **Time Series Data- Ordered data values observed over time**
- **Panel Data– Data observed over time from the same units of observation**

By Source: Primary or Secondary

- **Primary Data - data gathered for the first time by the researcher**
- **Secondary Data - Data taken by the researcher from secondary sources, internal or external o Already published records/compilation**

Sources of data

- **Primary source**

This refers to collecting data directly from the field. Such data, information collected by the population census enumerators, business survey enumerators, e.t.c.

- **Secondary Source**

This refers collecting data from published or unpublished compilations e.g. journals, newspapers, magazines, sales records, production records, textbooks e.t.c. Examples include:

- ✓ **Trade associations (e.g KACITA)**
- ✓ **Commercial services**
- ✓ **National and international institutions (e.g URA, UBOS, etc)**

Data Collection Considerations

Considerations to make before data collection

- Statement of the purpose
  - ✓ should be clearly be stated to avoid confusion
  - ✓ Only necessary information is collected
- Scope of inquiry
  - ✓ based on space or time- geographical and time
- Choice of statistical unit
- Data Sources
- Data collection technique
  - ✓ Depends on time available, literacy of the respondents, language, availability of the resources, the accuracy required

Advantages of primary data

- ✓ The data is original.
- ✓ The information obtained is unbiased.
- ✓ It provides accurate information and is more reliable.
- ✓ It gives a provision to the researcher to capture the changes occurring in the course of time.
- ✓ It is up to date, relevant and specific to the required product

Disadvantages of primary data

- ✓**Time consuming to collect**
- ✓**It requires skilled researchers in order to be collected.**
- ✓**It needs a big sample size in order to be accurate.**
- ✓**It’s more costly to collect**

Advantages of Secondary data

- ✓ It's economical. It saves expenses and efforts since it is obtainable from other sources.
- ✓ It is time saving, since it is more quickly obtainable than primary data.
- ✓ It provides a basis for comparison for data collected by the researcher.
- ✓ It helps to make the collection of primary data more specific, since with the help of primary data, one is able to identify the gaps and inefficiencies so that the additional or missing information may be collected

Disadvantages of Secondary data

- ✓ Accuracy of secondary data is not known.
- ✓ Data may be outdated.
- ✓ It may not fit in the framework of the research factors for example units used.
- ✓ Users of such data may not have as thorough understanding of the background as the original researcher.

Methods of data collections

- Observation,
- Interviews,
- Use of questionnaires,
- Use of mechanical devices,

Observation

- Uses own eyes to get information relevant to the research.
- **Advantages**
  - Subjective bias is minimized,
  - Information obtained is always most current / up to date.
  - Easy to handle/carry out, it does not require respondents.
- **Disadvantages**
  - Expensive,
  - Limited information is obtained,
  - Unforeseen factors can interfere with the process, not all things are visible.

Interview Method

- Researcher and a respondent come face to face
- Researcher asks questions as the respondent answers.
- Personal interview or a telephone interview, or Skype.
- Normally structured with pre-determined questions.
- **Advantages**
  - Realistic information can be obtained,
  - There is opportunity for probing questions allowing in depth details
  - Flexibility, easy to conduct and control.

Interview Method

- **Disadvantages**
  - Very expensive,
  - There may be some bias,
  - People are not easily approachable,
  - Time consuming.
  - Deliberate false information may be given.

Questionnaires

- An instrument to collect data in form of questions that are answered by the respondent and returned to the researcher
- Can be sent by mail or personally delivered
- Most commonly used by researchers in data collection.
- Questionnaires can be structured or unstructured.
- Structured questionnaires provide alternatives answers of which the respondents are expected to choose.
- Unstructured questionnaires, the questions are answered in the respondent's own words.

Questionnaires

- **Advantages**
  - Less costly,
  - Least biased
  - Gives respondents enough time to answer the questions,
  - Can cover a larger area.
- **Disadvantages**
  - Few questionnaires are returned,
  - Can only be used by the literate,
  - low flexibility in answering,
  - Can be very slow,

Developing a Questionnaire

- **Factors to consider when constructing a questionnaire**
  - Keep the research problem in ,
  - Keep questions simple and in line with the intended response and audience.
  - Check the sequence and formats of the questions.
  - Scrutinize and remove all technical defects.
  - Carry out a pilot study to test the questionnaire.

Selection of Data Collection Method

- When selecting a data collection method consider the following factors:
  - Type of inquiry to be carried out.
  - Available resources in terms of people and funds.
  - Time available to do the research.
  - The end user and nature of inference to be made.

Measurement scales

Measurement scales for Qualitative data

- Nominal scale
- Ordinal scale

Measurement scales for Quantitative data

- Interval scale-
- Ratio scale

Nominal scale

- It is used for variables that can be measured by classification only. Non-numerical in nature. **It involves only naming.**
- Categories without a meaningful order identify nominal data (Gender, political affiliation, industry classification, ethnic/cultural groups).

Examples of nominal scales

Sex of respondent	1=Male 2=Female
Religious affiliation	1= Anglican 2= Catholic 3= Born again2= Eastern 4= Moslem3= Northern 5= Others ...
Position held	1=Owner 2=Manager 3=Director 4=Other, Specify……

Ordinal scale

It involves ordering (its what's important and significant)

- It is a measurable scale which focuses or bases on ranking of ordered Categories.eg in athletics competition we have the first, second, third ……etc

Response scale	1= SD	2= D	3= N	4= A	5= SA
Tax Registration					
Tax officers are helpful to us when it comes to registering for taxes.					
We find it easy registering for taxes					
We do not lose so much time at registration for taxes					

Interval scale

- Interval scales are numeric scales in which we know not only the order, but also the exact differences between the values.
- An example of an interval scale is the Fahrenheit scale for measuring temperature i.e. the increments are known, consistent, and measurable.
- Interval scales are nice because the realm of statistical analysis on these data sets opens up. For example, *central tendency* can be measured by mode, median, or mean; standard deviation can also be calculated.

Interval scale-Continued

Like the others, you can remember the key points of an "interval scale" pretty easily. "Interval" itself means "space in between," which is the important thing to remember.

Interval scales not only tell us about order,

but also about the value between each item

The difference in Temperature between 100 degrees Centigrade and 90 degrees Centigrade is the same as 60 degrees Centigrade and 70 degrees Centigrade.

Ratio scale

Ratio scales are the best when it comes to measurement scales because they tell us about; the order, the exact value between units, AND they also have an absolute zero–which allows for a wide range of both descriptive and inferential statistics to be applied.

- Everything above about interval data applies to ratio scales + ratio scales have a clear definition of zero. Good examples of ratio variables include height and weight.

In Summary

Nominal scale is used to "*name*," or label a series of values.

Ordinal scales provide good information about the *order* of choices, such as in a customer satisfaction survey.

Interval scales give us the order of values + the ability to quantify *the difference between each one*.

Ratio scales give us the ultimate–order, interval values, plus the *ability to calculate ratios* since a "true zero" can be defined.

Data Analysis

- The process of trying to attach meaning to data.
- Initial stages of data analysis [Data Reduction]
  - *Data clean up* –look through the notice to remove undesirable data,
  - *Data editing* –here the researcher goes through the work editing it for consistency completeness and correctness.
  - *Field editing*. This done when collecting data.
  - *Central editing*. This is done at the central office where all data collected is looked through.
  - *Data sorting*. Sorting them according to the way data will be handled in the subsequent tasks.
  - *Data coding*. This involves the researcher placing codes on to the various categories formed during sorting.

Quantitative data analysis.

- Relates to quantifiable data eg figures, scales, weights, heights, and means.
- Can be Descriptive and inferential data analysis
  - Descriptive data analysis,
    - Descriptive statistics aims to summarize a sample.
    - Display using charts, diagrams, tables, frequencies and frequency distributions.
    - Mainly used when dealing with large samples of data and group observations.

Quantitative data analysis

- Inferential analysis
  - The data to learn about the population that the sample of data is thought to represent
  - Sample data used to derive conclusions on population.
  - Estimations of means, totals, and other attributes to describe relations and comparisons

Qualitative data analysis.

- Data that is unquantifiable, like opinion, perception, and attitudes.
- Follow three iterative steps
  1. Become familiar with the data through
     1. Reading
     2. Memoing
  2. Exam the data in depth to provide detailed descriptions of the setting, participants, and activities.
  3. Categorizing and coding pieces of data and grouping them into themes.
- Attach meaning to the results

Qual Vs Quan: Data collection method

	Quantitative	Qualitative
Sampling	Random sampling	Open ended and less structured protocols (Flexible)
Tools	Structured data collection instruments	Depend on interactive interviews
Results	Produce results that generalize, compare and summarize	Produce results that give meaning, experience and views

Questions or Comments

?