DATA IN A RESEARCH

Tran Thi Ut, FEC/HSU
October 10, 2013
IMPORTANCE OF DATA

• In research, it needs data for the stages:
  • Research design
  • Sampling design
  • Data gathering and/or field work techniques
  • Data processing and analysis
Distinguished Secondary data and Primary data

**Secondary data** is the data that have been already collected and recorded by someone else and readily available from other sources.

**Primary data** are gathered by researcher and usually done by survey research.
PURPOSE OF SECONDARY DATA

- Extracting the relevant information from other sources, previous studies
- **Fact findings**: Descriptive information to support research
- **Model Building**: specifying relationship between two or more variables
- **Data mining**: Exploring data through computer. Using computer technology to go through volumes of data to discover trends about an organization’s sales customers and products. IT is primarily used
- **Identifying the relevant sources** To avoid plagiarism
Advantages and Disadvantages of Secondary data

**Advantages**
- Faster
- Less Expensive
- Less activities (Field trip, Survey etc.)

**Disadvantages**
- Not easily available
- Not adequate
- May not meet the needs of researcher
- Outdated information
- Variation in definition
- Inaccurate or bias
## Evaluation of Secondary data

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
<th>Can the data be reworked</th>
<th>stop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do the data help to answer questions set out in the problem definition?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do the data apply to the time period of the interest?</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do the data apply to the population of the interest?</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do other terms and variable classifications presented</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are the units of measurement comparable?</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If possible go to the original source of the data?</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the cost of data acquisition worth it?</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is there a possibility of bias?</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can the accuracy of data collection be verified?</td>
<td>Yes</td>
<td></td>
<td></td>
<td>Stop</td>
</tr>
<tr>
<td>(accurate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If any question results in a 'No', the data may need to be reworked or discarded.
Primary data collection

- Qualitative data
- Quantitative data
Qualitative data types

- In-depth/unstructured interviews
- Semi-structured interviews
- Structured interview questionnaires containing substantial numbers of open comments
- Artifacts, pictures, drawings, physical objects
- Case study notes, videos, recordings
- Focus group notes, recordings, videos
- Action-oriented participation with the group being studied
- Participant notes, videos, recordings
- Unstructured or semi-structured diaries
- Minutes of meetings
- Personal documents (e.g. letters, personal diaries, correspondence)
- Press clippings
- Photographs or any other type of visual material
- Participant observation field notes, videos, recordings
- Technical fieldwork notes, etc.
- Kinship diagrams/other anthropological material
Types of qualitative data

- in-depth/unstructured interviews
- semi-structured interviews
- structured interview questionnaires containing substantial open comments
- focus group
- unstructured or semi-structured diaries
- observation field notes/technical fieldwork notes
- kinship diagrams/other anthropological material
- case study notes
- personal documents (e.g. letters, personal diaries, correspondence)
- press clippings
- photographs or any other type of visual material
Collection qualitative data

• Qualitative data: Many types - it is impossible to specify one ‘best’ way to collect data.
• flexibility about how qualitative researchers proceed with data collection, but not for sloppiness. (tùy tiện)
• Data collection must be carefully planned, executed, and controlled to gain scholarly respect.
Qualitative data collection method

Data collection approaches:

- Direct interaction with individuals on a one to one basis
- Direct interaction with individuals in a group setting
- It is time consuming, therefore data is usually collected from a smaller sample.
- The benefits: the information is richer and has a deeper insight into the phenomenon under study
- The main methods for collecting qualitative data are:
  - Individual interviews (in-depth interview)
  - Observations
  - Focus groups
  - Participatory approaches
In-Depth Interviewing

• Kahn and Cannell (1957) describe interviewing as “a conversation” with a purpose
• The participant’s views are valuable and useful
• Combined with observation, interviews allow the researcher to understand the meanings that everyday activities hold for people
• Listening skills and be skillful at personal interaction, question framing, and gentle probing for elaboration.
Observation:

*Observation is a fundamental and highly important method in all qualitative inquiry.*

- Systematic noting and recording of events, behaviors, and artifacts (objects) in the social setting chosen for study.

  • *field notes*—detailed,
  • nonjudgmental, concrete descriptions of what has been observed
OBSERVATION METHOD

Researcher can collect data the respondents through observation

- **Nonparticipant observer**: without becoming the integral part of the organization system
- **Participant observer**: Enters the organization, become a part of the work team
- **Structured observation studies**: Preparing number of areas to observe
- **Unstructured observation studies**: Researcher does not have define ideas of the particular aspects
Observation Method

Advantages:
• Data more reliable and free from respondent bias
• Easier to note the effects of environmental influences on the specific outcome
• Easier to observe certain groups of individual

Disadvantages:
• Observer must be present
• Low collecting data
• Needed to be trained
Focus Groups

• Using: Marketing research, social sciences and applied research.

• Group: 7 to 10 people, unfamiliar with one another

• The interviewer creates a supportive environment, asking focused questions to encourage discussion and the expression of differing opinions and points of view.

• These interviews may be conducted several times with different individuals
Disadvantages of focus group

• The researcher should be exquisitely aware of power dynamics and be able to facilitate well—
• Difficult control over a group
• Time can be lost;
• the data are difficult to analyze because context is essential to understanding the participants’ comments;
Participatory approaches

- PRA: Participatory Rural Appraisal
- PAR: Participatory Action Research
- Accessing the needs
- Feasibility study.
- Determination for priority activities in development study.
- Follow up development activities
Participatory Research (PR)

- PR is a systematic inquiry, with the collaboration of those affected by the issue being studied, for purposes of taking action or effecting change.
  
  *Green et al (2003:419)*

- PR: to break down the distinction between the researchers and the researched, the subjects and objects of knowledge production by the participation of the people-for-themselves.

- Research is seen not only as a process of creating knowledge, but simultaneously, as education and development of consciousness, and of mobilization for action.
Principles in participatory approach

(Israel et al (1998))

1. Recognizes community as a unit of identity.
2. Builds on strengths and resources within the community.
3. Facilitates collaborative partnerships in all phases of the research.
4. Integrates knowledge and action for mutual benefit of all partners.
5. Promotes a co-learning and empowering process that attends to social inequalities.
6. Disseminates findings and knowledge gained to all partners.
Focus Groups

• Using: Marketing research, social science and applied research.

• Group: 7 to 10 people, unfamiliar with one another.

• The interviewer creates a supportive environment, asking focused questions to encourage discussion and the expression of differing opinions and points of view.

• These interviews may be conducted several times with different individuals.
Participatory approaches

• PRA: Participatory Rural Appraisal
• PAR: Participatory Action Research
• Accessing the needs
• Feasibility study.
• Determination for priority activities in development study.
• Follow up development activities
Problem tree
Problem tree
Participatory Research (PR)

PR is defined as systematic inquiry, with the collaboration of those affected by the issue being studied, for purposes of education and taking action or effecting change. Green et al (2003:419)

PR: to break down the distinction between the researchers and the researched, the subjects and objects of knowledge production by the participation of the people for themselves.

Research is seen not only as a process of creating knowledge, but simultaneously, as education and development of consciousness, and of mobilization for action.
Maping
## Biến động theo thời gian

<table>
<thead>
<tr>
<th>NĂM</th>
<th>Sự Kiện</th>
<th>Ảnh hưởng đến</th>
<th>Lưu ý</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Trai giống thủy sản Thư Kêp</td>
<td>Giảm chất lượng</td>
<td>?</td>
</tr>
<tr>
<td>2006</td>
<td>Phân loại hải sản (rạm bảo)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>Tái thông tin P+ (để bắt ven bờ, đầy yếu)</td>
<td>Người lớn ven bờ</td>
<td>Giới hạn</td>
</tr>
<tr>
<td></td>
<td>15 m</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cải thiện hài bể</td>
<td>Khoi</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Phương Hài**

- Hỗ trợ đến 4
- Thiet hai bể
- Khích triplet

**29.03.2013 09:04**
Problem tree
Problem tree
Problem tree
Venn Diagram
Venn diagram
SWOT(C) analysis
<table>
<thead>
<tr>
<th>Điểm mạnh (S)</th>
<th>Cơ hội (O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Phát triển du lịch</td>
<td>- Vùng canh tác thủy sản của các tỉnh phía nam</td>
</tr>
<tr>
<td>- Phát triển du lịch biển</td>
<td>- Phát triển đô thị mới</td>
</tr>
<tr>
<td>- Giảm nghèo, tăng trungtwor</td>
<td>- Phát triển dịch vụ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Điểm yếu (W)</th>
<th>Cảm nhận (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Ngành nghề đơn giản</td>
<td>- Không có kinh nghiệm trong lĩnh vực</td>
</tr>
<tr>
<td>- Ngày đầu tiên trong năm</td>
<td>- Không có kinh nghiệm trong lĩnh vực</td>
</tr>
<tr>
<td>- Ngày đầu tiên trong năm</td>
<td>- Không có kinh nghiệm trong lĩnh vực</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(room number)</th>
<th>- Phản ứng với ngành kinh tế</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>(room number)</th>
<th>- Phản ứng từ người dân</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>(room number)</th>
<th>- Phản ứng từ người dân</th>
</tr>
</thead>
</table>
Quantitative data’

Survey

• A research technique for collecting primary data based on communication with respective sample of individuals/group of people

• **Survey objectives**: The type of information gathered depending on the survey ‘s objectives

• **Respondent**: The person who answers interview questions

• **Sample survey**: Representative sample of the target population

**Advantages of survey**

• Accurate results valuable

• Quick, Inexpensive, efficient accurate
Method of communication for collection primary data

- Personal interview
- Mail survey
- Tel. interview
- Self adm. Survey
Personal interview

Face to face situation:

**Advantages**
- Adapt the questions as necessary, clarify doubts, clearly understood, opportunity for feedback
- Obtain complete and precise information
- High participation rate
- Able to notice nonverbal cues from respondent

**Disadvantages:**
- Geographical limitation
- Cost of training interviewers
- Respondent might feel uneasy to sensitive issues
Telephone interview

Advantages:
• Big number of different people can be reached in short period of time
• Low cost
• Reduce/ eliminate discomfort of personal interview
• Respondents may feel less uncomfortable

Disadvantages:
• Length of interview is limited
• Refusing to answer
• Visual aids cannot be utilized
Structured and Disguise Questions

• Structured question limits the number of responses
• Unstructured question tend to be opened-ended, respondents feel freedom to answer
• The researcher can disguise the questions if the subject matter is of a threatening nature
ERRORS IN SURVEY

a) Random Sampling Error

b) Systematic Error: (non sampling method errors)
   - imperfect research design
   - mistake in execution of a research  ➔ Samples bias

The results of samples show a persistent tendency to deviate in one direction from the true value.

There are two general categories of systematic error:

• Respondent error: Respondents do not cooperate / do not give honest answers

• Administrative error – Data proceeding error, sample selection error, interviewer error
Systematic error

Respondent error

*Nonresponsive error*

Resp.- do not cooperate /do not give honest answers

a) Resp. did not respond to the questionnaires. Happened in mail survey/tel./personal survey

b) Response bias: Resp. tend to answer in a certain direction due to intentional or inadvertence or misrepresentation.

**Types of response bias:**

(i) Acquiescence bias: Say “yes”/”no” to all statement

(ii) Extremity bias

(iii) Interviewer bias: Interviewer influences the respondents answer

(iv) Social desirable bias: Respondent wanted to show prestige in the interviewer’s mind

Administrative error

Types

- Data processing error
- Sample selection error
- Interviewer error
- Interviewer cheating
Temporal Basis

a) Cross-sectional study:

• Data is collected at a single point in time.
• Various segments of the population are sampled. Relationships among variables may be investigated by cross-tabulation

b) Longitudinal study – Questioned at different points in time – changes occurring can be observed over time

• **Cohort study / tracking study** – Understanding several samples/groups at different times: Avoids bias
• Panel study: Gathering data from the sample over time to understand the trend
A sample: A subset or some part of a larger population representative of the entire population

A sample is drawn using either probability or non-probability sampling procedure

Non-probability sampling is typically applied in exploratory research. It is used when data needs to be collected speedily with little cost.

A population element: The individual participant or object to be measured
WHY SAMPLE?

Sampling is the process of using a smaller number of items /parts of a larger population to make conclusion about the entire population.

(a) Pragmatic Reasons (practical)

Research are carried out to solve problems in business entities and due to availability of limited (budget /time constraints)

(b) Provide Accurate and Reliable Results

Samples can provide sufficiently accurate findings and truly reflect the population if they are selected properly.

© Avoid Destruction of Test Units

Using samples for testing instead of testing all population would be avoid destruction of test units
SAMPLING DESIGN AND CONCEPTS

There are three main stages in sampling design and it should minimize the error that may occur due to the sampling process.

1: Should the study gather information from a sample or a census?
2: What type of sampling approach?
3: How about the appropriate size of the sample?
Stages in the selection of a sample

1. Define the Target Population
2. Select a Sampling Frame
3. Determine if a Probability or non-probability sampling
4. Plan Procedure for Selecting Sampling Units
5. Determine Sample size
6. Select Actual Sampling Units
7. Conduct Fieldwork
Defining the Target Population (TP)

TP: the complete group of objects or elements relevant to the research project.

Sampling unit: element/object available for selection during sampling process

Choosing the Sampling Frame: Comprehensive list of elements of sample:
- The telephone listing of restaurants or individual
- University’s registration list of all employees or customers
RANDOM SAMPLING ERROR AND NON-SAMPLING ERROR

Random sampling error: the difference between the sample result and the result of a census conducted using identical procedures.

Non-sampling error: results from some imperfect aspect of the research design due to response error or from a mistake in the execution of research.
TYPES OF SAMPLING

Probability Sampling
- Simple random Sampling

Probability of selection = \frac{\text{Sample size}}{\text{Population size}}

Population: 20,000 if sample size: 400, the probability of selection is 2\% \left(\frac{400}{20,000}\right)
Procedure for drawing large samples

i) Use Assign a unique identification number to each element in the sampling sequentially

ii) Use a random number generator to identify the appropriate elements to be selected

iii) Ensure that no element is selected more than once
Complex Probability Sampling

Simple random sampling is often impractical when:
Requires a population list (Sampling frame) that is not available
Fails to use all the information about a population
May be expensive to implement, in terms of time and money resources
Alternative probability sampling approaches:

*Systematic sampling*: every nth element in the population is sampled.

*Stratified sampling*: Divide the population to each stratum and get sample from that for increasing sample’s statistical efficiency.

*Cluster sampling*: To sample economically, getting sample from each cluster. Two conditions for use of cluster sampling:

  - The need for more economic efficiency than can be provided by simple random sampling.
  - The frequent unavailability of a practical sampling frame for individual elements.
## Comparisons of Stratified and Cluster Sampling

<table>
<thead>
<tr>
<th>Stratified Sampling</th>
<th>Cluster Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Divide population into few subgroups</td>
<td>1. Divide population into many subgroups</td>
</tr>
<tr>
<td>• Each subgroup has many elements</td>
<td>• Each subgroup has few elements</td>
</tr>
<tr>
<td>• Subgroups are selected according to some criteria related to variables</td>
<td>• Subgroups are selected according to some criterion of ease or availability in data collection</td>
</tr>
<tr>
<td>2. <strong>Try to secure homogeneity within subgroups</strong></td>
<td>2 Try to <strong>secure heterogeneity within subgroup</strong>.</td>
</tr>
<tr>
<td>3. <strong>Try to secure heterogeneity between subgroups</strong></td>
<td>3. <strong>Try to secure homogeneity between subgroups</strong></td>
</tr>
<tr>
<td>Randomly choose elements from within each group</td>
<td>4. Randomly choose several subgroups and study them in depth</td>
</tr>
</tbody>
</table>

1. **Try to secure homogeneity within subgroups**

2. **Try to secure heterogeneity within subgroup.**
Area Sampling

- Area sampling similar to cluster sampling base on geographical area.
- Applied to national populations, country population or even smaller area where there are well-defined political or natural boundaries

(d) Multistage Sampling: Combination of two or more other probability sampling techniques
- Progressively smaller geographical areas are selected in a series of steps until a housing unit is selected as the sample
Non-Probability Sampling

• Reasons studies use non-probability sampling
• When it satisfactorily meet sampling objectives
• There is no desire or need to generalize a population parameter
• Limited objectives
• Cost and time constraint
Non-Probability Sampling technique

• Convenience Sampling : in the case to have informal pools of friend and neighbor
• Test ideas or gain ideas about of subject of interest
• Using in early stage of exploratory research when She/he is seeking guidance or find out more problem on hand
• the population is unique
• The results may present evidence
Major alternative non-probability sampling techniques

- **Convenient sampling**: Collecting sample who are most conveniently available for the survey. (Economical/quickest way to obtain a large of responses)

- **Judgment Sampling**: Selecting sample members to conform to some criterion-Who have experiences to the problem

- **Quota Sampling**: This is another type of purposive sampling. Using to improve representativeness.

- To ensure the various sub-groups or categories in a population are represented in pertinent sample characteristics to the exact extent the investigators desire
Snowball Sampling (SS)

• SS is using in the case researcher has got difficulty to identify the best sample.
• Initial respondents are selected by probability samples; additional respondents are obtained by referral from initial respondents.
• Useful in locating members of rare populations.
• High bias because sample units are not independent, projecting data beyond samples is inappropriate.
Determination of Sample size

Factors are considered:

- Variability of elements in the target population
- Type of sample required
- Time available
- Budget
- The level of research estimation precision
- Whether the findings are to be generalized
Statistical Formula

Three major decisions must be made

The degree of confidence (Often 95%)
The specified level of precision (Amount of acceptable error)
The amount of variability of the population (Standard deviation)
Statistical Formula

\[ n = \left( \frac{zs}{E} \right)^2 \]

n = sample size

z = degree of confidence (the number of standard errors for the degree of confidence specified for the research

s = Variability (Standard deviation of the population)

E = Desired precision (The acceptable difference between the sample estimate and the population value)
Exp: Studying expenditures in lipstick, Confident level: 95% (z), a range of error (E) less than $2.00. The estimate of the standard deviation is $29.00.

\[
n = \left( \frac{z_s}{E} \right)^2 = \left[ \frac{(1.96)(29.00)}{2.00} \right]^2 = \left( \frac{56.82}{2.00} \right)^2 = (28.42)^2 = 808
\]
Given population size

Slovin (1960) : \( n = \frac{N}{1 + Ne^2} \)

\( \text{Trong đó } n: \text{Sample size} \)
\( N: \text{Population size} \)
\( e: \text{error (\%)} \)

Exp: \( N=9000, \text{Error 2\%}, \)
\[ n = \frac{9000}{1+9000(0.02)^2} \]
\[ n = \frac{9000}{1+9000(0.0004)} = \frac{9000}{1+3.6} = \frac{9000}{4.6} = 1.957 \]

Notes Assumption: Normal distribution
Sample size

Large population
No number of population

\[ n = \frac{z^2 \cdot (p \cdot q)}{e^2} \]

n= Sample size
z= value belong to the degree of error to be accept:

Exp:
90%,z =1.645
95% , z =1.96
99%, z = 2.327

p= estimated probability
q = 1-p

In general : p = 50% and q= 50%.

e = Error term to be accepted (+-3%, +-4%,+-5%...)
References
Israel, Barbara; Amy Schultz; Edith Parker and Adam Becker. 1998. *Review of Community-Based Research:*
Rohana Kamarudin, Noryyati Ahmad, 2010. *Business research method,* Open University Malaysia
Ut Tran Thi (2010) *Participatory Action Research,* Handout, Hong Bang University
VSO: *Sharing skills, changing live Participatory approaches:* A facilitator’s guide
Zikmund, w.g.(2003) *Business research methods.* Thomson South-Western, USA