Methodology - Sampling

Lecturer/Convenor:

Richard Boateng, PhD.

Email:
richard@pearlrichards.org

Office: UGBS RT18 (rooftop)
Learning Objectives

This session seeks to discuss the different approaches to sampling in research – probability and non-probability sampling.

• By the end of the session, students will be able to understand and explain sampling methods including snowballing, deviant sampling, random sampling, cluster sampling and purposive sampling.

www.vivafrica.net
Chapter 6

• Sampling


www.tinyurl.com/neuman2007
Sampling

• Sampling is a process of selecting samples from a group or population to become the foundation for estimating and predicting the outcome of the population as well as to detect the unknown piece of information.
Sampling Terminology

• Sample
• Population or universe
• Population element
• Census

Sample

• Subset of a larger population
• We make conclusions on a population by studying or investigating a sample

Population

• Any complete group of entities within which we want to explore, understand or predict a social phenomena
  – People
  – Cars

Census

- Investigation of all individual elements that make up a population

Sampling Frame

• A list of elements from which the sample may be drawn
  – Working population
  – Mailing lists - database marketers
  – List of students in year 2

Primary Goal

• Quantitative
  – To obtain a **representative sample** from the population
  – Generalize/predict findings on a population

• Qualitative
  – **Not about representativeness**, more focused on samples which enhance understanding
  – Collect cases, events, or actions that **clarify and deepen understanding** in a specific context

Neuman, W.L. (2011) Basics of Social Research: Qualitative and Quantitative Approaches, 2/E, Pearson Education
Two Major Categories of Sampling

• **Probability sampling**
  • Known, nonzero probability for every element

• **Nonprobability sampling**
  • Probability of selecting any particular member is unknown

Non-Probability Sampling

- This sampling technique is not based on random selection.
- Sample size is not determined in advance and the researcher has limited knowledge about the population from which the sample is being drawn.
  - Haphazard/Convenience sampling
  - Quota sampling
  - Snowballing
  - Purposive sampling
  - Deviant Sampling

Convenience/Haphazard Sampling

• The sampling procedure of obtaining the people or units that are most conveniently available
• Cases are obtained in any manner which is convenient but high possibility of being ineffective
• Can produce highly unrepresentative samples
  – Person on street interviewed for TV
  – Cut-out a newspaper questionnaire and mail it in

Quota Sampling

• Ensures that the various subgroups in a population are represented on **pertinent** sample characteristics to the **exact extent** that the investigators desire
  – Identify the relevant categories (e.g., gender and age)
  – Set a quota for each category

• It should not be confused with stratified sampling.

• Better than haphazard sampling

Judgment/Purposive Sampling

• An experienced individual selects the sample based on his or her judgment about some appropriate characteristics required of the sample. Often used in a exploratory research

• Selecting particular cases for in-depth investigation

• Selecting members difficult to reach
  – Research on prostitutes

• Selecting unique cases which are informative
  – Studying failed development project
  – Selecting the popular trend setting women magazine for a content analysis study

Snowball Sampling

• Network, chain referral or reputational sampling
• Identifying samples in a network
  – multistage – beginning with a few people and grow through referral
• Initial respondents are selected by other methods like purposive sampling or random sampling
• Additional respondents are obtained from information provided by the initial respondents

Deviant Sampling

- Searching cases that differ from the dominant pattern
- Use various techniques to identify cases with specific characteristics that differ from the dominant
  - School dropouts who seem not to have no record of illegal activities and who are stable from two-parent, upper-middle income families

Neuman, W.L. (2011) Basics of Social Research: Qualitative and Quantitative Approaches, 2/E, Pearson Education
Probability Sampling

• **Advantages**
  – Saving time and cost
  – Accuracy

• **Types**
  – Simple random sampling
  – Systematic sampling
  – Stratified sampling
  – Cluster sampling
Simple Random Sampling

• A sampling procedure that ensures that each element in the population will have an equal chance of being included in the sample
• Uses mathematical theory to select elements
• Random-Number Table
  – For a population = 1000
    – 1st selection = 1:1000
    – 2nd selection = 1:999
## Random Number Table

### Extract from Random Number Table

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

### Sample Selected

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>02</td>
<td>Adwoa</td>
</tr>
<tr>
<td>11</td>
<td>Kay</td>
</tr>
<tr>
<td>04</td>
<td>Grace</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Kwame</td>
</tr>
<tr>
<td>02</td>
<td>Adwoa</td>
</tr>
<tr>
<td>03</td>
<td>Nan</td>
</tr>
<tr>
<td>04</td>
<td>Grace</td>
</tr>
<tr>
<td>05</td>
<td>Hui</td>
</tr>
<tr>
<td>06</td>
<td>Pael</td>
</tr>
<tr>
<td>07</td>
<td>Nisci</td>
</tr>
<tr>
<td>08</td>
<td>Eren</td>
</tr>
<tr>
<td>09</td>
<td>Renee</td>
</tr>
<tr>
<td>10</td>
<td>Siade</td>
</tr>
<tr>
<td>11</td>
<td>Kay</td>
</tr>
</tbody>
</table>
Systematic Sampling

• Every \( k \text{th} \) name/element from the list will be drawn.
• You select the first element /name at random and the subsequent element /name by Sample interval.
• **Sample Interval** is standard distance between elements selected in a sample. It is cyclical.
• Sample Interval = \( \frac{\text{population size}}{\text{sample size}} \)
• Sample ratio = \( \frac{\text{sample size}}{\text{population size}} \)
• **Sample ratio** is the proportion of elements in the population that are selected.

Neuman, W.L. (2011) Basics of Social Research: Qualitative and Quantitative Approaches, 2/E, Pearson Education
Systematic Sampling

- **Starting number:**
  The researcher selects an integer that must be less than the total number of individuals in the population. This integer will correspond to the first subject.

- **Interval:**
  The researcher picks another integer which will serve as the constant difference between any two consecutive numbers in the progression. The integer is typically selected so that the researcher obtains the correct sample size.

Read more: [http://www.experiment-resources.com/systematic-sampling.html#ixzz28gK1Kn92](http://www.experiment-resources.com/systematic-sampling.html#ixzz28gK1Kn92)
Systematic Sampling - Example

• From the sampling frame, a starting point is chosen at random, and thereafter at regular intervals.
• For example, suppose you want to sample 8 students from a database of 120 students.
• \( \frac{120}{8} = 15 \), so every 15th student is chosen after a random starting point between 1 and 15. If the random starting point is 11, then the students selected are 11, 26, 41, 56, 71, 86, 101, and 116.
Systematic Sampling of Cyclical Data

This table presents the population of students in an English class.

You are required to use systematic sampling to select a sample. Your sampling interval is 2. Randomly starting from number 9 (Renee), which student will be the last to be selected?

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Kwame</td>
</tr>
<tr>
<td>02</td>
<td>Adwoa</td>
</tr>
<tr>
<td>03</td>
<td>Nan</td>
</tr>
<tr>
<td>04</td>
<td>Grace</td>
</tr>
<tr>
<td>05</td>
<td>Hui</td>
</tr>
<tr>
<td>06</td>
<td>Pael</td>
</tr>
<tr>
<td>07</td>
<td>Nisci</td>
</tr>
<tr>
<td>08</td>
<td>Eren</td>
</tr>
<tr>
<td>09</td>
<td>Renee</td>
</tr>
<tr>
<td>10</td>
<td>Siade</td>
</tr>
<tr>
<td>11</td>
<td>Kay</td>
</tr>
<tr>
<td>12</td>
<td>John</td>
</tr>
</tbody>
</table>
Systematic Sampling of Cyclical Data

This table presents the population of students in an English class.

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Kwame</td>
</tr>
<tr>
<td>02</td>
<td>Adwoa</td>
</tr>
<tr>
<td>03</td>
<td>Nan</td>
</tr>
<tr>
<td>04</td>
<td>Grace</td>
</tr>
<tr>
<td>05</td>
<td>Hui</td>
</tr>
<tr>
<td>06</td>
<td>Pael</td>
</tr>
<tr>
<td>07</td>
<td>Nisci</td>
</tr>
<tr>
<td>08</td>
<td>Eren</td>
</tr>
<tr>
<td>09</td>
<td>Renee</td>
</tr>
<tr>
<td>10</td>
<td>Siade</td>
</tr>
<tr>
<td>11</td>
<td>Kay</td>
</tr>
<tr>
<td>12</td>
<td>John</td>
</tr>
</tbody>
</table>

Your sample size = 12 / 2 = 6

First to be selected is Renee

2\textsuperscript{nd} – Kay
3\textsuperscript{rd} – Kwame
4\textsuperscript{th} – Nan
5\textsuperscript{th} – Hui
6\textsuperscript{th} - Nisci

Other literature ignore the first randomly selected and choose by sample interval to obtain the first.
Interval (fractional/decimal)

• For example, suppose you want to sample 8 students from a database of 125 students.

• If there were 125 students, $\frac{125}{8}=15.625$, so should you take every 15th student or every 16th student? If you take every 16th student, $8 \times 16=128$ so there is a risk that the last student chosen does not exist. To overcome this the random starting point should be between 1 and 13.

• On the other hand if you take every 15th house, $8 \times 15=120$ so the last five houses will never be selected. The random starting point should now be between 1 and 20 to ensure that every house has some chance of being selected.

• Population - ((Sample size – 1) * Sample Interval)) = Extreme starting point
Systematic Sampling

• **Advantages**
  – spreads the sample more evenly over the population
  – easier to conduct than a simple random sample

• **Disadvantage**
  – the system may interact with some hidden pattern in the population
Stratified Sampling

• Subsamples are drawn within different strata (distinct groups)
• Each stratum is more or less equal on some characteristic. Do not confuse with quota sample
  – In study of 20,000 students you decide to pick 200 students. However, you are informed that 2 percent of the 20,000 students are foreign students. You need that representation in your 200 students.
  – Hence, you collect the list of the foreign students (400) and randomly select 4 students (2 % of 200) to include in your sample 200 students.

Cluster Sampling

• The purpose of cluster sampling is to sample economically while retaining the characteristics of a probability sample.

• The primary sampling unit is no longer the individual element in the population. The primary sampling unit is a larger cluster of elements located in proximity to one another.

  – We need to study a population of 1,000 students concerning banking services preferences. Our objective is to interview 400 students. However, the students are divided into 10 different programmes of study, each enrolling a minimum of 90 students. In this case, we can randomly select 4/5 clusters which add up to 400.


Neuman, W.L. (2011) Basics of Social Research: Qualitative and Quantitative Approaches, 2/E, Pearson Education
## Examples of Clusters

<table>
<thead>
<tr>
<th>Population Element</th>
<th>Possible Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>College seniors</td>
<td>Colleges</td>
</tr>
<tr>
<td>Airline travelers</td>
<td>Airports Planes</td>
</tr>
<tr>
<td>Sports fans</td>
<td>Football stadiums Basketball arenas Baseball parks</td>
</tr>
</tbody>
</table>
Systematic Errors

- Errors which systematically affect the measurement of the variable across the whole sample
  - Unrepresentative sample results
  - Not due to chance; Due to study design or imperfections in execution
  - For instance, if there is loud traffic going by just outside of a classroom where students are taking a test, this noise is liable to affect all of the children's scores -- in this case, systematically lowering them.
  - Unlike random error, **systematic errors tend to be consistently either positive or negative** -- because of this, systematic error is sometimes considered to be **bias** in measurement.

Random Error

• Random error is caused by any factors that randomly affect measurement of the variable across the sample.
• For instance, each person's mood can inflate or deflate their performance on any occasion. In a particular testing, some children may be feeling in a good mood and others may be depressed. If mood affects their performance on the measure, it may artificially inflate the observed scores for some children and artificially deflate them for others. The important thing about random error is that it does not have any consistent effects across the entire sample. Instead, it pushes observed scores up or down randomly.
• The important property of random error is that it adds variability to the data but does not affect average performance for the group. Because of this, random error is sometimes considered noise.

Errors Associated with Sampling

• Sampling frame error
  – Sampling frame is biased

• Nonresponse error
  – Failure to adequately collect data on variables


5. OUM (2010) Principles Supporting Qualitative Research, Topic 9 Qualitative Research Methods, Course Hand out CMRM6103 Research Methodology/GMRM5103 Research Methods, AIT Open University of Malaysia, Ghana


Lecturer/Convenor:

Richard Boateng, PhD.

Email:
richard@pearlrichards.org

Office: UGBS RT18 (rooftop)

Questionnaire Design and Levels of Measurements
Level of Measurement

• Variables is anything that can take a numerical value

• **Continuous variables** have an infinite number of values or attributes that flow along a continuum – temperature, age, income, and crime rate.

• **Discrete variables** have a relatively fixed set of separate values or variable attributes – gender (male or female), marital (never married single, married, divorced or separated, widowed)
Types of Variables

Extraneous Variables

Variables that affect the relationship

Independent Variables

CAUSE
Change Variables

Variables that link

Intervening Variables

Dependent Variables

EFFECT
Outcome Variables

Types of variables in causal relationship
(Adapted from Research Method: Ranjit Kumar, SAGE Publications, page 60)

OUM (2010) Principles Supporting Qualitative Research, Topic 9 Qualitative Research Methods, Course Hand out CMRM6103 Research Methodology/GMRM5103 Research Methods, AIT Open University of Malaysia, Ghana
Types of Variables

• change variables are referred to as **independent** variables while outcome variables are known as **dependent** variables. On the other hand, the unmeasured variables affecting the cause-effect relationship are called **extraneous** variables and the variables that link a cause and effect linking are called **intervening** variables.

OUM (2010) Principles Supporting Qualitative Research, Topic 9 Qualitative Research Methods, Course Handout CMRM6103 Research Methodology/GMRM5103 Research Methods, AIT Open University of Malaysia, Ghana
IT Awareness in a Rural Community

Independent, dependent and extraneous variables

OUM (2010) Principles Supporting Qualitative Research, Topic 9 Qualitative Research Methods, Course Hand out CMRM6103 Research Methodology/GMRM5103 Research Methods, AIT Open University of Malaysia, Ghana
Level of Measurement

• Variables is anything that can take a numerical value

1. Conceptualization
   – Continuous (temperature – degrees)
   – Discrete (temperature – cold, hot)

2. Type of indicator
   - Nominal – indicate that there is a difference among categories; e.g. Gender – male and female
   - Ordinal – indicate that there is a difference among categories and the categories can be ordered (e.g. Grade: A, B, C, D)
   - Interval – measures everything nominal and ordinal do, plus they specify the amount of distance between categories. This measurement has a starting and a terminating point that is divided into space intervals (Celsius scale; income intervals: $100-$249; $250-$399).
   - Ratio – measures everything all the others do, plus there is a true zero, which makes it possible to state relations in terms of proportion or ratios (e.g., money income: $0, $100, $500 or years of formal schooling: 1 year, 10 years, 13 years)
Level of Measurement

• Variables is anything that can take a numerical value

– Continuous Variables
  • Can be expressed as Interval or Ratio

– Discrete
  • Can be expressed as Ordinal or Nominal
Questionnaire Design

• Questionnaires need a logical structure

• Well-thought-out structure

Pilot Survey

• A pilot survey is generally a small-scale run through of the survey and can also be used to check questionnaire coding and method of analysis.
Question Structure

Five possible objectives of a question

a) To find if the respondent is aware of the issue
   - Do you know of any plans to build a school in this community?

b) To get general feelings on an issue
   - Do you think a school should be built?
     ▪ A rating scale can be used for this type of question

c) To get answers on specific parts of the issue
   - Do you think a school will affect the local environment?

d) To get reasons for a respondents views
   - Why are you against the motorway being built?

e) To find how strongly these views are held
   - How important is the tourist center that would be demolished if the school is built?
Question Coding

• Precoded questions give the respondent a series of possible answers from which one may be chosen or an alternative specified.
  – How many children do you have?
    • 0 1 2 3 4 5 6 7

• Sometimes codes are developed from the answers.
  – Where do you live?

• An open question will allow the respondent to say whatever he or she wishes:
  – Why do you choose to live in Kumasi?
Question wording - bias

• Two or more questions presented as one
  – *Do you use self-service garages because they are easy to use and clean?*
  – YES/NO

• Questions that contain difficult or unfamiliar words
  – *Where do you usually shop?*
  – How often is usual? Shopping also vary in terms of type of product, day of week and time of the year
Question wording bias

• Questions which start with words meant to soften hardness or directness
  – *I hope you don’t mind me asking this, but are you a virgin?*
  – YES/NO

• Questions which contain conditional or hypothetical clauses
  – *How do you think your life would change if you had nine children?*
  – This is a situation that few people will have considered….

• Questions which contain one or more instructions to respondents
  – *If you take your weekly income, after tax, and when you have made allowances for all of the regular bills, how much do you have left to spend or save?*
References

3. OUM (2010) Principles Supporting Qualitative Research, Topic 9 Qualitative Research Methods, Course Hand out CMRM6103 Research Methodology/GMRM5103 Research Methods, AIT Open University of Malaysia, Ghana