Introduction

- Strength of quantitative research method – its ability to use smaller group of people to make inferences about larger groups (Bartlett, Kotrik & Higgins, 2001).
  - Refers to: making generalization findings from sample back to the population.
- To do so – you need to pick the most accurate smaller group to represent the larger group. [This is called the Sampling]
  - This smaller group = SAMPLE (n)
  - The larger group = population (N)
**Definitions: Population vs Sample**

- **What is a sample?**
  - A finite part of a statistical population chosen to be studied to represent the population.
  - (Symbol = \( n \))

- **What is a population?**
  - A population is a group of individuals, persons, objects, or items from which samples are taken for measurement.
  - Data collected from the whole population – census.
  - (Symbol = \( N \))

**Sampling**

- **What is sampling?**
  - Sampling is a technique of selecting a suitable sample, or a representative part of a population for the purpose of determining parameters or characteristics of the whole population using a range of methods.

- Therefore in Sampling technique:
  - Select sample from population.
  - There are several methods/techniques to choose the sample.
Important considerations in Sampling

- What is the most appropriate sampling technique to use?
  - Does my sample represent the population?
  - What is my sampling frame

- What size sample do I need?

Sample Techniques

- **Probability Sampling**
  - Random Sampling
  - Systematic Random Sampling
  - Stratified Random Sampling
  - Cluster Sampling
  - Stage Sampling

- **Non Probability Sampling**
  - Purposive Sampling
  - Quota Sampling
  - Convenience Sampling
  - Snowball Sampling
Probability Sampling

- **Probability Sampling**
  - The chance/probability of each case being selected from the sample.
  - Allows to make inference from sample about population (generalization)
  - Example:
    Sample = Consumer. Rate the price of Chocolate bar. 75% said expensive.
    Inference: 75% of all consumer feels the same.
- Use of **inferential statistics** – the Significance value (p-value) and confidence interval.

Types of Probability Sampling Techniques

- **Simple Random Sampling (Sample Rawak Mudah)**
  - Select sample at random from sampling frame
  - How:
    - Number each cases in your sampling frame with a unique number. 1,2,3,……
    - Select cases using random numbers until you get the desired sample size.
    - Cast lots or computer program – Excel, SPSS (if data available in computer).
  - Example: Population – 200 Desired Sample: 100
  - Advantage: Easy to implement
  - Disadvantage: Require list of population elements, over or under representativeness, time-consuming
Random Sampling using SPSS
### Systematic Random Sampling – Sample rawak sistematik

- Select sample at regular intervals from the sampling frame.
- How:
  - Number each cases in sampling frame with unique number. 1,2,3,4,.....
  - Select the 1\(^{st}\) case using a random number.
  - Calculate sampling fraction
    - Actual sample size/total population: 100/500 = 1/5 (select one in every 5 case)
  - Select subsequent cases (from the 1\(^{st}\) case) using the sampling fraction.

#### Sampling Frame Data

<table>
<thead>
<tr>
<th>Case</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Advantages
• Simple to design
• Easier than simple random

Disadvantages
• Periodicity within population may skew sample and results
• Trends in list may bias results
• Moderate cost

Stratified Random Sampling
• Divide the population according to several strata (sub-sets).
• A random or systematic sample is drawn from the sub-sets/strata. How.
  • Choose the stratification variable (eg: age group, job status, gender, etc.)
  • Divide the sampling frame into the sub-set
  • Number each of the cases within each subset with a unique number.
  • Select sample using simple or systematic sampling proportionately or disproportionately.
Example: Choosing employees using stratified sampling technique

Stratified Sampling

M = Male
FM = Female
What is proportionate and disproportionate stratified sampling?

**Proportionate**
- The number of elements from each stratum is selected according to its proportion in the population.
- Example:
  - Total population $N = 500$ (Male = 300, Female = 200).
  - Proportion of male & female:
    - Male = $300/500 \times 100 = 60\%$
    - Female = $200/500 \times 100 = 40\%$
  - Sample size $n = 350$. Proportion of male & female in the sample:
    - Male ($n$) = $350 \times 60\% = 210$
    - Female ($n$) = $350 \times 40\% = 140$

**Disproportionate**
- The number of elements from each stratum is selected without consideration to the size of the stratum.
- Example:
  - Total population $N = 500$ comprise of Male & Female.
  - Sample size ($n$) = 350.
    - Select 50% male & 50% female.
      - Male ($n$) = $350 \times 50\% = 210$
      - Female ($n$) = $350 \times 50\% = 140$

**Advantages**
- Control of sample size in strata.
- Increased statistical efficiency
- Provides data to represent and analyze subgroups
- Enables use of different methods in strata

**Disadvantages**
- Increased error if subgroups are selected at different rates
- Expensive if strata on population must be created
- High cost
Cluster Sampling (Persampelan Kluster)

- Divide the population into discrete groups.
- The complete lists of the clusters (not individual/each elements) will serve as the sampling frame.
- Select a few clusters using Simple Random Sampling.

How:
- Choose the cluster grouping in your sampling frame.
- Number each of the clusters with a unique number. 1,2,3,4,.....
- Select the Cluster using Simple Random Sampling.
- Collect data from the population of the subset.

- Two conditions foster the use of cluster sampling.
  - the need for more economic efficiency and
  - The unavailability of a practical sampling frame for individual elements.

Advantages
- Provides an unbiased estimate of population parameters if properly done
- Economically more efficient than simple random
- Lowest cost per sample
- Easy to do without list

Disadvantages
- Often lower statistical efficiency due to subgroups being homogeneous rather than heterogeneous
- Moderate cost
Stage Cluster Sampling

- An extension of cluster sampling that involves successive random selection. How:
  - Choose the cluster grouping in your sampling frame.
  - Number each of the clusters with a unique number. 1,2,3,4,.....
  - Select the cluster using Simple Random Sampling.
  - Then randomly select certain proportion from the cluster as your sample.
- Advantages/Disadvantages: As in Slide20

Things to consider in sample selection using Probability Sampling.

- Identify correct Population and Sampling Frame (complete list of all cases in the population)
- Ensure representativeness of the sample (avoid sampling error).
- Avoid sampling bias (avoid sampling bias).
- Adequate sample size.
- High response rate.
  - Total response rate = total number of responses/total number in sample – (ineligible + unreachable).
Sample Representativeness

Sample Representativeness – not well represented
Estimating Sample Size:

<table>
<thead>
<tr>
<th>Required Sample Size</th>
<th>Confidence Levels</th>
<th>Margin of Error</th>
<th>Confidence Levels</th>
<th>Margin of Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Size</td>
<td>90%</td>
<td>10%</td>
<td>95%</td>
<td>5%</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>23</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>50</td>
<td>31</td>
<td>50</td>
<td>31</td>
<td>50</td>
</tr>
<tr>
<td>100</td>
<td>43</td>
<td>60</td>
<td>43</td>
<td>60</td>
</tr>
<tr>
<td>250</td>
<td>57</td>
<td>80</td>
<td>57</td>
<td>80</td>
</tr>
<tr>
<td>500</td>
<td>69</td>
<td>100</td>
<td>69</td>
<td>100</td>
</tr>
<tr>
<td>1000</td>
<td>81</td>
<td>120</td>
<td>81</td>
<td>120</td>
</tr>
<tr>
<td>2000</td>
<td>89</td>
<td>140</td>
<td>89</td>
<td>140</td>
</tr>
<tr>
<td>5000</td>
<td>97</td>
<td>160</td>
<td>97</td>
<td>160</td>
</tr>
<tr>
<td>10000</td>
<td>99</td>
<td>180</td>
<td>99</td>
<td>180</td>
</tr>
</tbody>
</table>

This formula is the one used by Krejcie & Morgan in their 1970 article “Determining Sample Size for Research Activities” (Educational and Psychological Measurement, #30, pp. 607-610).

This document is a sample of a population, well represented.
• Formula for calculating sample size depending on type of statistical analysis:
  • Regression/Correlation analysis (Tabachnik & Fidell, 2001):
    • \( n \geq 104 + m \) (independent variable) – Simple linear regression.
    • \( n \geq 50 + 8m \) (independent variable) – Multiple linear regression.

Non-Probability Sampling

• The probability for each case being selected from the total population is not known.
• Cannot make inference from the sample about the population. Cannot make GENERALIZATION
• Most often used in qualitative studies.
• In some quantitative studies it may not be possible to use probability sampling.
Non-probability Sampling Methods

- Convenience sampling – select sample purely on the basis that they are available.
- Snowball sampling – researcher identifies a small number of subjects who in turn identifies others in the population.
- Quota sampling – the researcher non-randomly selects subjects from identified strata until the planned number of subjects is reached.
- Purposive sampling – researcher deliberately selects the subjects against one or more trait to be a representative sample.

Cont.......
Steps in Sampling Design

- What is the target population?
- What are the parameters of interest?
- What is the sampling frame?
- What is the appropriate sampling method?
- What size sample is needed?

Procedure for Drawing a Sample

1. Step 1: Define the population
2. Step 2: Identify the sampling frame – listing of all units in the population from which the sample will be selected
3. Step 3: Select a sampling procedure
4. Step 4: Determine the sample size
5. Step 5: Select the sample units
6. Step 6: Collect data from the sampled units