Sampling of Populations

1

What is Sampling?

- Process of selecting a number of study units from a defined study population.
- Act of studying or examining only a segment of population to represent the whole.

2

Questions considered in sampling

- What is the group of people (study population) we are interested in from which we want to draw a sample?
- How many people do we need in our sample?
- How will these people be selected?

3

Advantages of sampling

- Lower cost
- Shorter time
- Better quality of information
- More comprehensive data may be obtained
- Only possible method for destructive procedures

4

Definition of Terms

- Population – is the universe about which an investigator wishes to draw conclusions.
- Sample – is a subset of population – the part that is actually being observed or studied.
- Element – is a single observation
- Probability samples – samples in which researcher can specify the probability of any one element in the population being included

5

Definition of Terms

- Target population – is the group from which representative information is desired and to which inferences will be made
- Sampling population – is the population from which a sample will actually be taken.
- Sampling unit – is the units which are chosen in selecting the sample, and may be made-up of non-overlapping collection of elements.

6

Definition of Terms

- Sampling frame – collection of all the sampling units
- Sampling error – is the difference between the population value of parameter being investigated and the estimates of this value based on the different samples

7

Criteria of Good Sampling Design

- Representativeness – reflect both characteristics and variability of population being studied.
- Sample size adequate
- Practicality and feasibility of sampling procedure
- Economy and efficiency of sampling design

8

Sampling

- Study population
  - Persons
  - Villages
  - Institutions
  - Records
  - Etc…

9

Study population consists of Study Units.

Defining both depends on the problem investigated and the objectives of the study

10

Representativeness

- Draw conclusions that are valid for the whole study population.
- Has all the important characteristics of the population from which it is drawn.

11

Basic Sampling Designs

- Depends on:
  - Nature of the variables
  - Population being studied
  - Purpose of research undertaken
  - Availability of information relevant to sampling procedure

12

Sampling Designs Types

- Probability
- Non-probability

13

Non-Probability Sampling

- Probability that each member of population to be selected in sample is difficult to determine or cannot be specified.
- No way to assess reliability of sample results
- Standard errors cannot be computed
- Methods of statistical inference cannot be applied

14

Non-Probability Sampling

- Best for descriptive purposes
- Not good for making generalizations or inferences about target population.

15

Non-probability Sampling

- Purposive sampling strategies for qualitative studies
  - Focus on limited number of informants
  - Respondents are selected strategically so that their in-depth information will give optimal insight into an issue where little is known.
  - Depend on topic under study
  - Depend on type of information wanted
  - Depend on resources of the investigator(s)

16

Purposive Sampling

- Extreme case sampling
  - Selection of extreme cases like good or very poor compliers to treatment, AIDS patients
- Maximum variation sampling
  - Researcher wants to obtain as complete as possible insight to a certain issue in all its variations.
  - Example: Determining the stigma of leprosy, HIV
  - Use quota sampling (specific groups and interviews a fixed number per group)

17

Purposive Sampling

- Homogeneous sampling
  - Obtaining specific information about 1 particular group only.
  - Common: FGD
- Typical case sampling
  - Illustrative in describing in-depth some cases which are 'typical' for the group one is interested in.
  - Example: Describing typical family in community X, typical problems of miners, etc.
  - Cannot generalize for the whole group
  - Typical examples can be selected with cooperation of key informants who know the study well, from a survey & the characteristics investigator is interested in.

18

Purposive Sampling

- Critical case sampling
  - Critical cases are those who ‘can make the difference’ with respect to an intervention you want to introduce or to evaluate.
- Snowball or chain sampling
  - Suitable in locating key informants or critical cases.
  - In-depth interview leads to discoveries.

19

Purposive Sampling is NOT same as Convenience Sampling

20

Convenience Sampling

- For convenience sake.
- Study units happen to be available at time of data collection are selected in the sample.
- When no other choice (no one else available) researchers may use this approach.
- Haphazard or accidental sampling

21

Quota Sampling

- Data collectors are given quotas to meet
- Pre-specified quota

22

Probability sampling

- Rules and procedures for selecting sample and estimating parameters are explicitly and rigidly specified.
- Each unit in population has a known, non-zero chance of being included in sample

23

Random Sampling Strategies

- Collect quantitative data
- Measure variables distributed in a population
- Generalize the findings obtained from a sample to the total study population.

24

Probability Sampling

- Use random selection procedures to ensure that each unit of the sample is chosen on the basis of chance.
- All units of the study population should have an equal or at least a known chance of being included in the sample.
- Requires a listing of all study units (sampling frame) exists or can be compiled.

25

Probability Sampling Methods

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Cluster sampling
- Multistage sampling

26

Simple Random Sampling

- Every element has equal chance of being included in sample
- Simplest form
- Random numbers
- Sampling frame or a listing

27

Using random numbers

- Assume a sample of 10 households desired from a community of 80 households
- Since the number 80 represent population size, has 2 digits, then random numbers will be taken two at a time.
- Specific numbers to choose in table depends on the random start

28

Random Numbers

64249	63664	39652	40646
26538	44249	04050	48174
05645	00512	76630	55328
74897	68373	67359	51014
20872	54670	35017	68132

29

Sampling Frame

- Listing
- School record
- Map
- Hospital record
- Etc…

30

Systematic Sampling

- Regular intervals from sampling frame.
- Sampling fraction/ interval:

$$\frac{\text{Sample Size}}{\text{Study population}} = \text{Sampling fraction/ interval}$$

- Less time consuming & easier to perform than simple random sampling.
- Risk of bias, systematic variation

31

Stratified Sampling

- Small groups or STRATA
- Possible when we know what proportion of the study population belongs to each group we are interested in.
- Can take a relatively large sample from a small group in a study population.

32

Cluster Sampling

- Selection of groups of study units (clusters)
- Clusters are often geographic units or organizational units

33

Steps in cluster sampling

- Population is first divided into sampling units and a sample of such units is selected
- Every element found in each sampling unit drawn as sample will be included in the study

34

Multi-stage Sampling

- Very large and diverse populations may be done in 2 or more stages
- Often in community-based studies
- Carried out in phases and usually involve more than 1 sampling method
- Advantages: Sampling frame not required, Sample is easier to select
- Disadvantage: May not be representative to total study population

35

Steps in Multi-stage sampling

- Population is first divided into a set of primary or first-stage sampling units. A sample of such units is selected.
- Each primary sampling unit included in the sample is further subdivided into secondary or second-stage sampling units from which a sample will again be taken.
- The procedure continues until desired stage is reached.

36

Combination of Designs

37

Bias in Sampling

- Systematic error in sampling procedures leads to a distortion of results
- Can be introduced as consequence of:
  - Improper sampling procedures
  - Non-response

38

Dealing with Bias

- Data collection tools should be pre-tested.
- Follow-up of non-respondents
- Separate study of non-respondents may be considered to identify to what extent they differ from respondents.
- Include additional people in the sample.

* Bigger non-response rate, more necessary to take remedial action.

39

Other Sources of Bias

- Studying volunteers only – they may be different from the study population.
- Sampling of registered patients only –may differ systematically from people seeking alternative treatments.
- Missing cases of short duration
- Seasonal bias
- Tarmac bias – may differ systematically from more inaccessible areas.

40

Generalizations on Estimation of Sample Size

- Determine sample size early if possible during the design phase.
- Formula for sample size determination depends on nature of outcome or parameters estimated, type of research design and sampling design used.
- Total population size
- Rare condition
- Complex data analysis

41

Generalizations on Estimation of Sample Size

- More heterogeneous the values of parameter of interest, larger required sample size
- Longitudinal studies require large sample size than case-control and cross-sectional studies
- Provide a correction factor for non-response or refusal rate in the estimation of sample size.
- Cluster sampling designs require greater sample size than simple random sampling

42

Generalizations on Estimation of Sample Size

- Higher level of accuracy and precision, larger sample size.
- Sample sizes are estimated separately for each important item or parameter
- Estimating for different units of analysis, sample sizes has to be estimated for each unit.

43

How to determine sample sizes?

- Qualitative studies
- Quantitative studies

44

Sample size in qualitative studies

- No fixed rules
- Size depends on WHAT you try to find out, from what different informants or perspectives you try to find that out.

- Note: Sampling procedures and sample size should be taken with care to avoid the allusion of haphazardness.

45

Sample size in quantitative studies

- Calculations
- Bigger sample is better .
  - Not true.
  - Increase the accuracy and richness of data collection (improving the quality of data)
  - Give a better representative samples

46

Rules to determine sample size

- Desired sample size depends on expected variation in the data (esp. of most important variables)
  - More varied data, larger sample size → to attain desired level of accuracy
- Desirable sample size depends on number of cells in cross-tabulations to analyze results.
  - Rough guideline: At least 5 to 10 study units per cell

47

Note

- Sample size is usually compromised between what is desirable and what is feasible

48

Sample size calculations

- Depends on whether the study:
  - Seeks to measure 1 single variable (e.g. mean, rate, proportion) in 1 group with a certain precision.
  - Demonstrate a significant difference between two groups.

49

Point estimates & their Standard errors

Measurement Scale	Population Parameter	Point estimate	Standard Error
Qualitative	Proportion (P)	Sample proportion, p	$(pq/n)^{1/2}$
Qualitative	Difference between 2 proportions (P1 – P2)	p1 – p2	$[p1q1/n1 + p2q2/n2]^{1/2}$
Quantitative	Mean, μ	Sample mean, X	$\delta/\sqrt{n}$
Quantitative	Difference between 2 means, μ1 – μ2	X1 – X2	$[(\delta_1^2/n1) + (\delta_2^2/n2)]^{1/2}$

50

Measuring 1 variable

- Single mean

  $n = s^2/e^2$

  n = sample size

  s = standard deviation

  e = required size of standard error ('margin of error' is used for $\pm$ 2 times the size of standard error if a precision of 95% is required)

51

Measuring 1 variable

- Single rate

  $n = r/e^2$

  r =rate

52

Measuring 1 variable

- Single proportion

$n = pq/e^2$

$q = (100 - p)$

53

Difference between 2 means

$n = s_1^2 + s_2^2 / e^2$

54

Difference between 2 rates

$n = r^1 + r^2 / e^2$

55

Difference between 2 proportions

$n = p_1q_1 + p_2q_2 / e^2$

56

Significant difference between 2 groups

- Comparison of two means

$n = (u + v)^2 (s_1^2 + s_2^2)/ (m1 – m2)^2$

u = one-sided % point of normal distribution, corresponding to 100% - the power. Power is probability of finding a significant result

v = % point of normal distribution corresponding to 2-sided significance level.

57

Significant difference between 2 groups

- Comparison of two rates

$n = (u + v)^2 (r_1^2 + r_2^2)/ (r1 – r2)^2$

u = one-sided % point of normal distribution, corresponding to 100% - the power. Power is probability of finding a significant result

v = % point of normal distribution corresponding to 2-sided significance level.

58

Significant difference between 2 groups

- Comparison of two proportions

$n = (u + v)^2 (p_1q_1 + p_2q_2)/ (p1 – p2)^2$

u = one-sided % point of normal distribution, corresponding to 100% - the power. Power is probability of finding a significant result

v = % point of normal distribution corresponding to 2-sided significance level.

59

Sample Size

- Tables for calculating sample size
- Statistical softwares.

60