Lesson: Sampling

Research Methodology - COMC/CMOE/ COMT 41543

By
Srinath Dissanayake
DCFM

Email: neelacksha911@gmail.com
Web: http://srilankatalkforum.weebly.com/index.html

1

Introduction..

- Surveys are useful and powerful in finding answers to research questions through ***data collection*** and ***subsequent analysis***, but they can do more harm than good if ***the population is not correctly targeted.***
- *Definition:* The process of selecting the right individuals, objects, or events as ***representatives for the entire population*** is known as sampling.
- *Definition:* Sampling is the process of selecting a ***small number of elements*** from ***a larger defined target group*** (Population) of elements such that the information gathered from the small group will allow ***judgments*** to be made about the larger groups.

2

The need to sample

- For some research questions it is possible to collect data from an *entire population* as it is of a *favorable size*.
- Sampling provide a valid alternative to a census when;
  - It would be *impractical* for you to survey the entire population.
  - Your *budget constraints* prevents you from surveying the population.
  - Your *time constraints* prevents you from surveying the population.

3

Sampling terminology

- **Population:** the population refers to the ***entire group of people, events, or things*** of interest that the researcher wishes to investigate.
- **Element:** An element is a single member of the population.
- **Sample:** a sample is a subset of the population.
- **Subject:** is a single member of the sample, *(just as an element is a single member of the population)*
- **Population parameters or population mean:** from sample statistics we make an estimate of the answers to our research questions in the study population. Estimates arrived at from **sample statistics** are called as **population parameters**.

4

Principles of sampling

- **Principle 01:** in a majority of cases of sampling there will be a ***difference*** between the sample statistics and the true population mean.
- Select a sample of two individuals to make an ***estimate*** of the average age of the four individuals.

1. A+B=18+20=38/2=19 years
2. A+C=18+23=41/2=20.5 years
3. A+D=18+25=43/2=21.5 years
4. B+C=20+23=43/2=21.5 years
5. B+D=20+25=45/2=22.5 years
6. C+D=23+25=48/2=24 years

5

Principles of sampling *Cont'd..*

Sample	Sample average age (1) (sample statistics)	Population mean (2) (population parameter)	Difference between (1) & (2)
1	19	21.5	-2.5
2	20.5	21.5	-1.5
3	21.5	21.5	0.0
4	21.5	21.5	0.0
5	22.5	21.5	+1.0
6	24	21.5	+2.5

Principles of sampling *Cont'd..*

- **Principle 02:** The **greater the sample size**, the **more accurate** will be the estimate of the true population mean.
- Let's Select a sample of three individuals to make an ***estimate*** of the average age of the four individuals.

1. A+B+C=18+20+23=61/3=20.33 years
2. A+B+D=18+20+25=63/3=21.00 years
3. A+C+D= 18+23+25=66/3=22.00 years
4. B+C+D=20+23+25=68/3=22.67 years

7

Principles of sampling *Contd..*

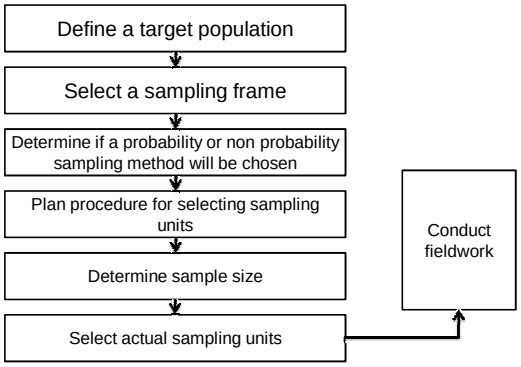
Sample	Sample aver age (1) (sample statistics)	Population mean (2) (population parameter)	Difference between (1) & (2)
1	20.33	21.5	-1.17
2	21	21.5	-0.5
3	22	21.5	+0.5
4	22.67	21.5	+1.17

Principles of sampling *Contd..*

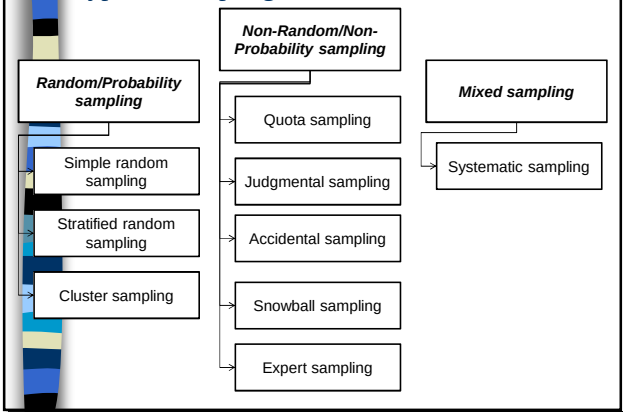
- ***Principle 03:*** The **greater the difference** in the variable under study in a population for a given sample size, the greater the will be the difference between the sample statistics and the true population mean.
- Suppose the ages of four individuals are markedly different: A=18, B=26, C=32, D=40.
- Eventually you will find that the difference in the average age in the case of samples of **two** ranges between -7.00 and +7.00 years and in the case of the sample of **three** ranges between -3.67 and +3.67.

9

Stages in selecting a sample



Types of sampling



Random /Probability sampling designs

- In random/probability sampling, it is imperative that each element in the population has an ***equal*** and ***independent*** chance of selection in the sample.
- Methods of drawing a random sample
  - The fishbowl draw
  - Computer programs
  - A table of randomly generated numbers:

12

Random /Probability sampling designs

- **Simple Random Sampling (SRS):** this is in line with the definition of randomization. Which is each element of the population is given an **equal** and **independent** chance of selection.
- **Stratified Random Sampling:** Involves dividing the population into **mutually exclusive** and mutually exhaustive subgroups /strata and then taking a ***simple random sample*** in each subgroup or strata.
- *Example:* subgroups - sex, age group, religion or geographical regions.
- It is important to that the characteristics chosen as the basis of stratification are **clearly identifiable** in the study population.
- E.g:. It is easier to stratify the population on the basis of gender than age.

13

Random /Probability sampling designs

- **Stratified Random Sampling**
  - **Proportionate Stratified Random Sampling:** The number of elements from each stratum in relation to its *proportion in the total sample* is selected.
  - **Disproportionate Stratified Random Sampling:** Consideration is not given to the *size* of the stratum.
- **Cluster Sampling**: Cluster sampling is more appropriate where, the desired **population is large.**
- Cluster sampling is based on the ability of the researcher to divide the sampling into groups, called clusters, and then to select elements within each cluster, using the SRS technique.
- Initially, the researcher needs to divide the population into clusters (usually along geographical regions). Then he randomly select some clusters from all clusters selected.

14

Non-Random /Non-Probability sampling designs

- Non-Probability sampling designs do not follow the ***theory of probability*** in the choice of elements from the sampling population.
- Non-Probability sampling is used when the number of elements in a population is either ***unknown or cannot be individually identified.***
- **Quota Sampling:** the main consideration in quota is the ***ease of assessing the sample population.***
- And also the researcher is guided by some ***visible characteristic***, such as gender or race, of the study of the population that is of interest to him/her.
- Sample is selected from ***a location convenient to the researcher***.
- The process continues until the researcher has been able to contact the required number of respondents (quota).
- Example !!

15

Non-Random /Non-Probability sampling designs

- **Accidental Sampling:** not like in Quota sampling, accidental sampling is ***not guided by any visible characteristic.***
- ***Example:*** ***in a market survey some people contacted may not have the required information.***
- **Judgmental or Purposive Sampling: here the *judgment* is important. (i.e. *who can provide the best information*)**
- **Expert Sampling:** judgmental sampling was based on judgment, whereas, expert sampling must be ***best known experts in the field of interest to you.***
- **Snowball Sampling:**
- **Snowball sampling attempts to select a sample using a *network.***
- ***To start with, a few individuals in a group or organization are selected and the required information is collected from them. They asked to identify other people in the group or organization, and the people selected by them become a part of the sample.***

16

Mixed sampling designs

- **Systematic sampling design:** this has the both characteristics of ***random*** as well as ***non random*** sampling designs.

- ***Steps***
- 1st, prepare a list of all the elements in the study population (N)
- 2nd, decide on the sampling size (n)
- 3rd, determine the width of the interval (k)
- Width of the interval = $\frac{\text{Total population}}{\text{Sample size}}$
- 4th, using the SRS, select an element from the first interval (nth order).
- 5th, Select the same order element from each subsequent interval.

17

Mixed sampling designs

- ***Example:***

Sampling frame

1	Interval	14	26	39
2		15	27	40
3		16	28	41
4		17	29	42
5		18	30	43
6		19	31	44
7		20	32	45
8		21	33	46
9		22	34	47
10		23	35	48
11		24	36	49
12		25	37	50
13			38	

Sample selected

3
8
13
18
23
28
33
38
43
48

The calculation of sample size

- The ***size of the sample*** is important for testing hypothesis or establishing an association.
- The method depends on whether the estimation;

$$\text{Sample Size} = \frac{PQ\,(Z^2)}{E^2}$$

- Proportion of absentees = 15%
- Q=(1-P)=100-15=85
- Z= Given confidence interval; 1.96 (if 95% C.I)
- E= Maximum deviation from true proportion that can be tolerated. (say 5%)
- If P=15%; Z=95% confidence interval; Proportion (E) 5%

$$\text{Sample Size} = \frac{15 \times 85 \times 1.96^2}{5^2} = 195.9$$

19

References

- Kumar, R., 2011. *Research Methodology - A step by step guide for beginners*. 3rd ed. New Delhi: Sage Publications India Pvt Ltd.
- Sekaran, U. & Bougie, R., 2010. *Research Methods for Business: A skill building approach*. 5th ed. Delhi: Wiley India Publishers.

- Acknowledgements: Professor W. Indralal De Silva, Senior Professor of Demograhy (Chair), Department of Demography, Faculty of Arts, University of Colombo.

20