Study Population and Sampling

By
Dr. Eng. Dickinson Dunstan Turinawe

Population

- Population (Universe)
  - A group of individuals, events or objects having a common observable characteristic
  - Each member of a population is called a population element.
- Example
  - All P7 students in Uganda
  - All indigenous trees in National Forests
  - All students of MUBS
- Census -Investigation of all elements in a population
- A Researcher first defines the population to which He or She wants to generalize the results. This is the target population or Universe

Sampling

- A Sample is a small sub-group obtained from the population. Each member of the sample is referred to as a subject
- Why sample
  - Time and budget constraints
  - Accessibility to elements
- Sampling is the process of selecting number of elements for a study in such a way that the individuals selected represent the larger group.
- Sampling frame – A list of elements from which a sample may be drawn. Also known as working population

eg a list of students from the Registrar's office at a University.

Population vs Sampling Frame- Population is general while a sampling frame is specific.

Sampling designs/ sampling techniques

- **A sampling design is a process by which units of the population are drawn into the sample.**
- **The selection of the units can either be based on probabilities or no probabilities involved hence resulting into two broad sampling designs;-**
- **Probability and non sampling techniques**
- **For probability sampling, each unit in the study population a same chance of being included into the sample.**
- **It is also referred to as random sampling or sampling without bias.**
- **Under non probability sampling, there is no randomness in the selection of sampling units into the sample.**
- **Sampling is not based on chance. It is therefore also referred to as non random sampling or sampling with bias.**

A probability sampling

- Sampling method that utilizes some form of *random selection*. Most common method is use of random numbers:
- To generate random numbers use Command =INT(N*RAND())+1
- Random selection method
  - *Simple random sampling*. Here all elements have equal chance to be representative, and all elements are part of the population
  - *Systematic random sampling*. Here the sample is chosen systematically by choosing at random the first and then every n^{th} element.
  - *Stratified random sampling*. Here the population is divided into strata's using either proportionate or non proportionate stratification

$$\frac{\text{No of elements in the strata}}{\text{Total population}} \times \text{the required sample}$$

A probability sampling (Cont'd)

- Cluster Sampling
  - Identify the population
  - Define clusters forming the population
  - Determine the required sample size
  - List clusters in a random order
  - Select randomly the number of clusters depending on sample size
  - All members in the cluster are included in the sample as units of observation
    - Example: Schools, towns, hospitals Government Ministries

A probability sampling (Cont'd)

- Single-stage cluster sampling vs two-stage cluster sampling,
  - Limits generalization
- *Area sampling*. This relates to geographical locations. When cluster sampling is based on geographical sub divisions, it's is known as area sampling. E.g. if you are to consider districts from the east, to represent the whole country.
- *Multi stage sampling*. Here we use more than one sampling technique to come up with a more reliable sample.

SAMPLING FRAME

- Sampling Frame

A list of all items in your population. A population is general while a sampling frame is specific.

- population is too large to access directly;
- perhaps some elements of the population are more difficult to locate
- there are numerous pragmatic problems that arise in sampling populations
- Sampling frames must be assessed for all the above features, but
- particularly for completeness and potential bias

SAMPLE SIZE

- Sampling frame error- Occurs when certain sample elements are excluded or when the entire population is not accurately represented in the sampling frame.
- Sample Size

The Cochran formula is: $n=\frac{Z^2 p(1-p)}{d^2}$ for N>10,000

p= proportion of the target population estimated to have characteristic being measured

d= the level of statistical significance

Z= the standard normal deviation at the required significance

EXAMPLE

Suppose we are doing a study on the inhabitants of a large town, and want to find out how many households serve breakfast in the mornings. We don't have much information on the subject to begin with, so we're going to assume that half of the families serve breakfast: this gives us maximum variability. So p = 0.5. Now let's say we want 95% confidence, and at least 5 percent—plus or minus—precision. A 95 % confidence level gives us Z values of 1.96, per the normal tables, so we get

$((1.96)^2 (0.5) (0.5)) / (0.05)^2 = 385.$

So a random sample of 385 households in our target population should be enough to give us the confidence levels we need.

SAMPLE SIZE (Con'd)

For small populations and for N<10,000

$$n_f = \frac{n}{1 + \frac{n-1}{N}}$$

Here n is Cochran's sample size recommendation, N is the population size, and n is the new, adjusted sample size. In our earlier example, if there were just 1000 households in the target population, we would calculate

385 / (1 + (384 / 1000)) = 278

SAMPLE SIZE (Con'd)

- **Yamane's Formula**

$$n = \frac{N}{1+N(e)^2}$$

Consider a population of 5,000

N= 5000

e = 0.05

$$n = \frac{5000}{1+5000(0.05)^2} = \frac{5000}{1+12.5} = 370$$

SAMPLE SIZE (Con'd)

- Krejcie & Morgan (1970) came up with a table for determining sample size for a given population for easy reference

SAMPLE SIZE (Con'd)

Krejcie & Morgan (1970)

Table 3.1

Table for Determining Sample Size of a Known Population

N	S	N	S	N	S	N	S	N	S
10	10	100	80	280	162	800	260	2800	338
15	14	110	86	290	165	850	265	3000	341
20	19	120	92	300	169	900	269	3500	346
25	24	130	97	320	175	950	274	4000	351
30	28	140	103	340	181	1000	278	4500	354
35	32	150	108	360	186	1100	285	5000	357
40	36	160	113	380	191	1200	291	6000	361
45	40	170	118	400	196	1300	297	7000	364
50	44	180	123	420	201	1400	302	8000	367
55	48	190	127	440	205	1500	306	9000	368
60	52	200	132	460	210	1600	310	10000	370
65	56	210	136	480	214	1700	313	15000	375
70	59	220	140	500	217	1800	317	20000	377
75	63	230	144	550	226	1900	320	30000	379
80	66	240	148	600	234	2000	322	40000	380
85	70	250	152	650	242	2200	327	50000	381
90	73	260	155	700	248	2400	331	75000	382
95	76	270	159	750	254	2600	335	1000000	384

Note: N is Population Size; S is Sample Size *Source: Krejcie & Morgan, 1970*

STRATA NUMBERS

TM	50	100(50/500)
MM	150	100x(150/500)
W	300	100x(300/500)
TOTAL	500	100

Non probability sampling

- Used when approaching the sampling problem with a specific plan in mind.
  - Convenience (accidental),
  - Judgmental
  - Purposive sampling.
  - Snowballing

Non probability sampling (Biased Sampling)

- Convenience (accidental)
  - Selection of sample based on convenience to the researcher
- Judgmental
  - Sample selection based on researchers own knowledge, skills and experience
- Purposive sampling
  - Focuses on the right respondents of the research process

Questions or Comments

?