Sampling

Question ?

- Who appears to constitute the population of interest?
- Which type of sampling procedure best describes that used by the Researcher?
- What are the limitations of this sampling method, and in what specific ways could the sampling method have affected the findings?
- Why is it so important to get the size of a sample as close as possible to what is "correct"?

Key Terms in Sampling

- Population
- Target population
- Sample
- Sampling Frame
- Sampling
- Statistics
- Parameter
- Sampling error
- Generalizability
- Biased sample
- Sample Design

Population

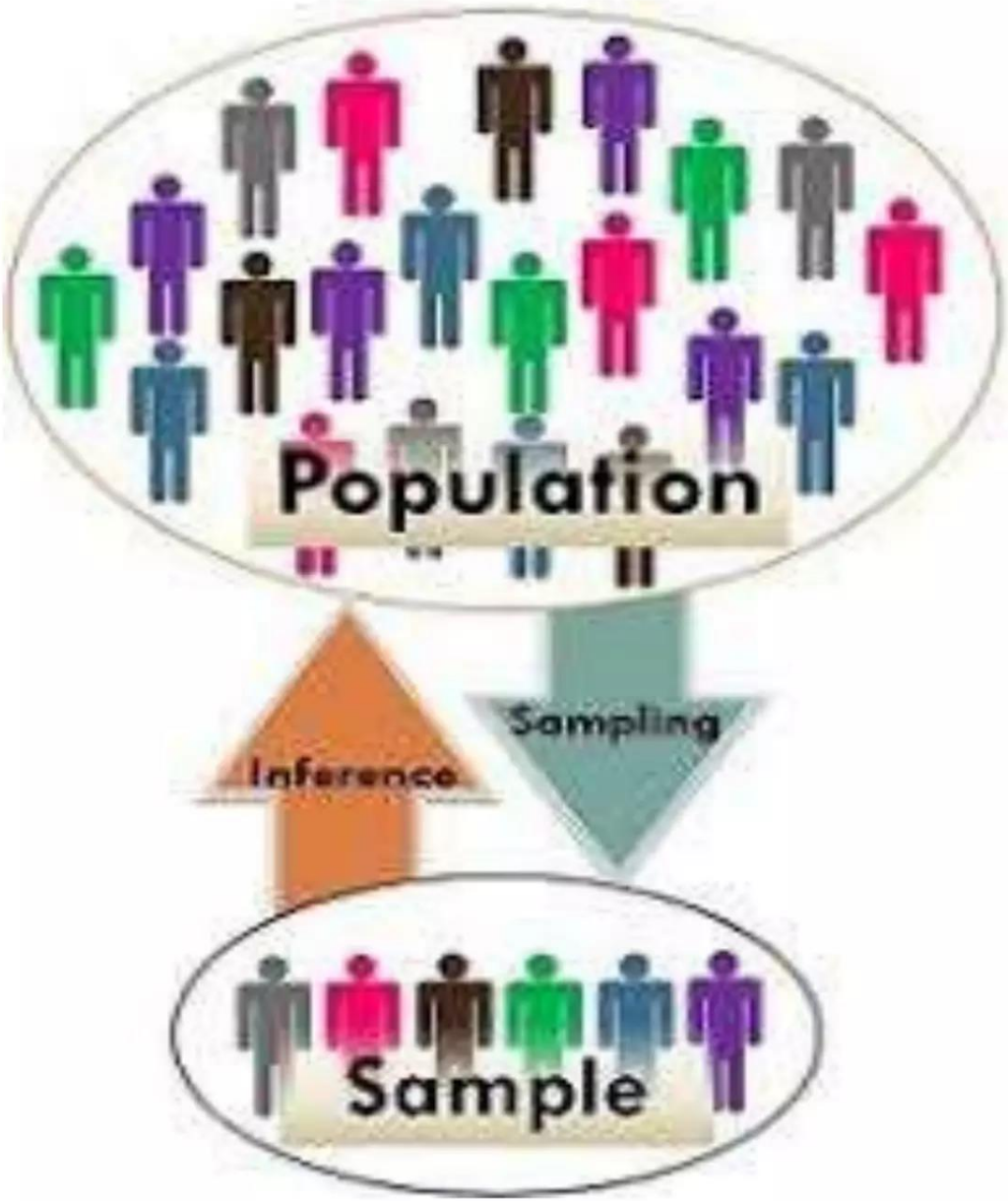
- In statistics, a population is the entire group of individuals, objects, or events that a study is interested in.
- A population is a complete set of persons or objects that possess some common characteristics that is of interest to the researcher.
- It's also known as the target population or research population.
- The population is defined by the research objectives and the attributes being investigated.
- It refers to the entire group of individuals or objects to which researchers are interested in generalizing the conclusions
- Example –
- JNU Ph.D. students Research Methodology Understanding
- Hospital patients

Target Population

- Target Population represents specific subgroup within the population that a researcher is interested in researching and analysing.
  - Students at a particular University -JNU
  - Of a particular programme- Economics, Sociology, History
  - Patient of disease- TB, Diabetes, Heart
  - Employees of a particular sector –Formal and Informal Sector
- The target population consists of people or things that meet the designated set of criteria of interest to the researcher.
- The group of individuals or items the researcher want to apply their study results generally is know as the target population.
- Target population depends on Research Objectives

Sample

- A sample is a **subset** of the population.
- The concept of sample arises from the inability of the researchers to test all the individuals in a given population.
- The sample must be representative of the population from which it was drawn and it must have good size to undergo statistical analysis.


Population
Sampling
Inference
Sample

Sampling frame

- A sampling frame is a list, database, or a source of information that contains all the elements or members of the target population from which a sample can be drawn.
- It serves as the actual source from which you select your sample.
- Ideally, the sampling frame should include every member of the target population, ensuring that each individual has a chance of being selected in the sample.
- It is a list of all those within a population who can be sampled, and may include individuals, households or institutions.

1. **Customer Satisfaction Survey**

- **Scenario:** A company wants to assess customer satisfaction for its new product.
- **Target Population:** All customers who purchased the new product in the last three months.
- **Sampling Frame:** The company's sales database listing all customers who made a purchase in the last three months, including their contact details.
- **Explanation:** The sales database serves as the sampling frame, as it contains the relevant subset of the customer base that the company wants to survey.

2.Academic Research on Student Behavior

- **Scenario:** A university researcher wants to study the online learning behavior of students.
- **Target Population:** All undergraduate students enrolled in the university.
- **Sampling Frame:** The university's student registration records, which include all currently enrolled undergraduate students.
- **Explanation:** The student registration records provide a comprehensive list of all potential participants, ensuring that the sample can be drawn from all undergraduate students.

3.Health Study on Elderly Population

- **Scenario:** A health department wants to study the prevalence of hypertension among elderly residents in a city.
- **Target Population:** All residents aged 60 and above in the city.
- **Sampling Frame:** The city’s healthcare database or voter registration list that includes information on residents' ages.
- **Explanation:** The healthcare database or voter registration list acts as the sampling frame, allowing the researchers to identify and sample individuals who are 60 years and older.

4.A health study on diabetes prevalence among adults in a city

- **Population:** All adults in the city.
- **Target Population:** Adults aged 40-60 in the city.
- **Sample:** 500 adults aged 40-60 selected from the target population.
- **Sampling Frame:** Voter registration lists or healthcare records.

- **Sampling frame vs. population**

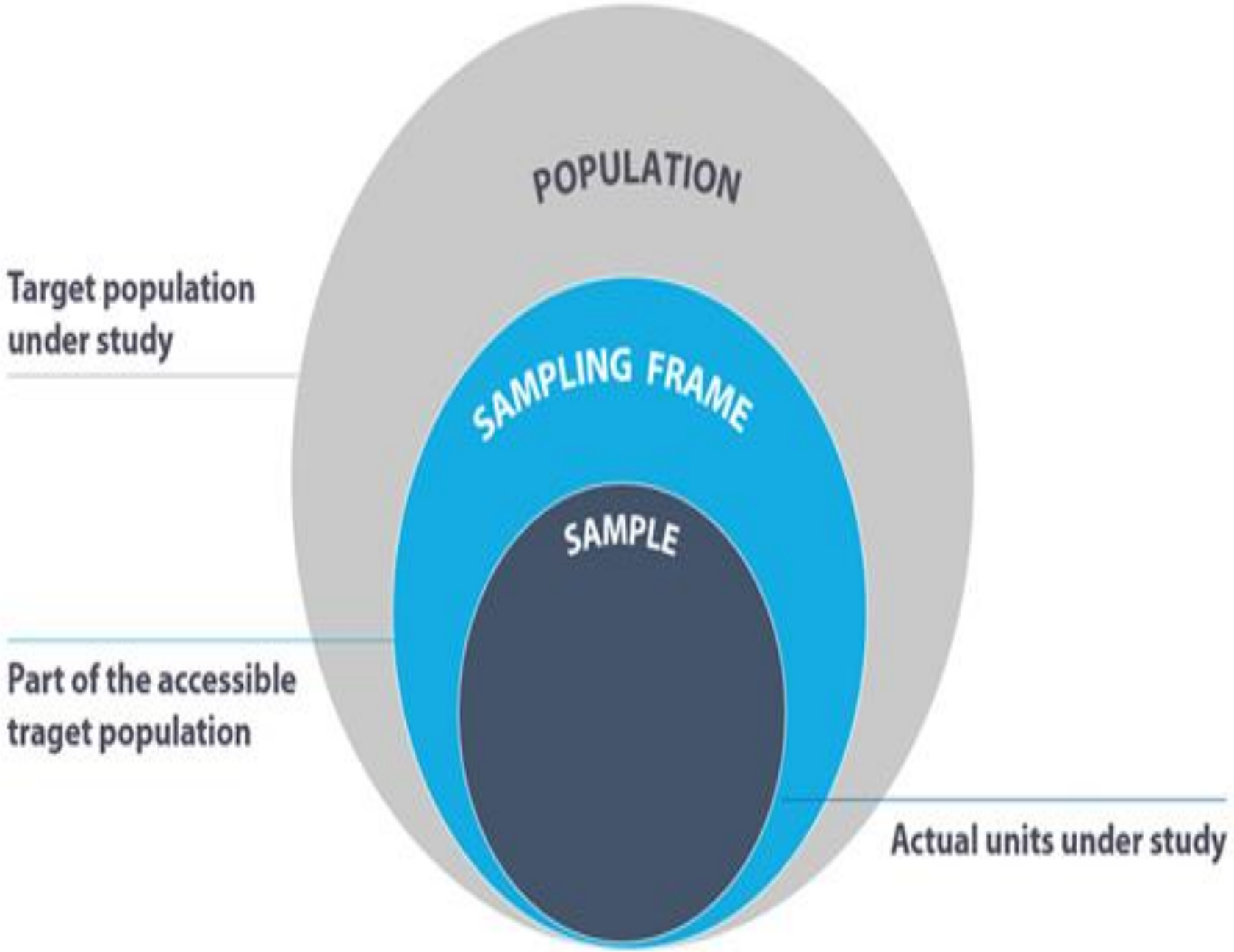
- The sampling frame *is a list of all people* or *units from the population that the sample* is selected from. In contrast, the population is *the whole group* or *collection of humans, objects, or events* that are being observed.

Sampling frame vs. target population

The sampling frame, *a portion of the target population*, is a *reference for researchers as they decide on and choose the proper sample*. It describes the list of people, homes, and other things that researchers have access to or can utilize to select a sample from. *The group of individuals or items the researchers want to apply their study results generally is known as the target population*

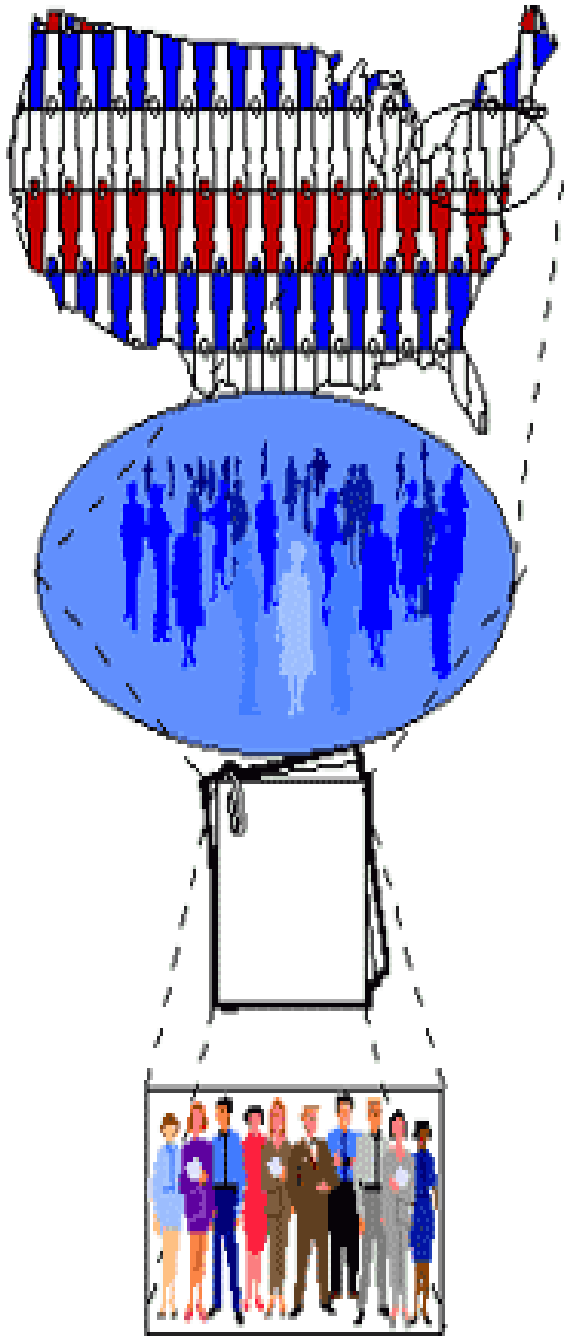
Sampling

- Sampling is simply the process of learning about the population on the basis of sample drawn from it (population). Therefore, in sampling , only part of the population is studied and the conclusions are drawn on that basis for the entire population or universe .


POPULATION
Target population under study
SAMPLING FRAME
SAMPLE
Part of the accessible traget population
Actual units under study

Who do you want to generalize to?

The Theoretical Population



What population can you get access to?

The Study Population

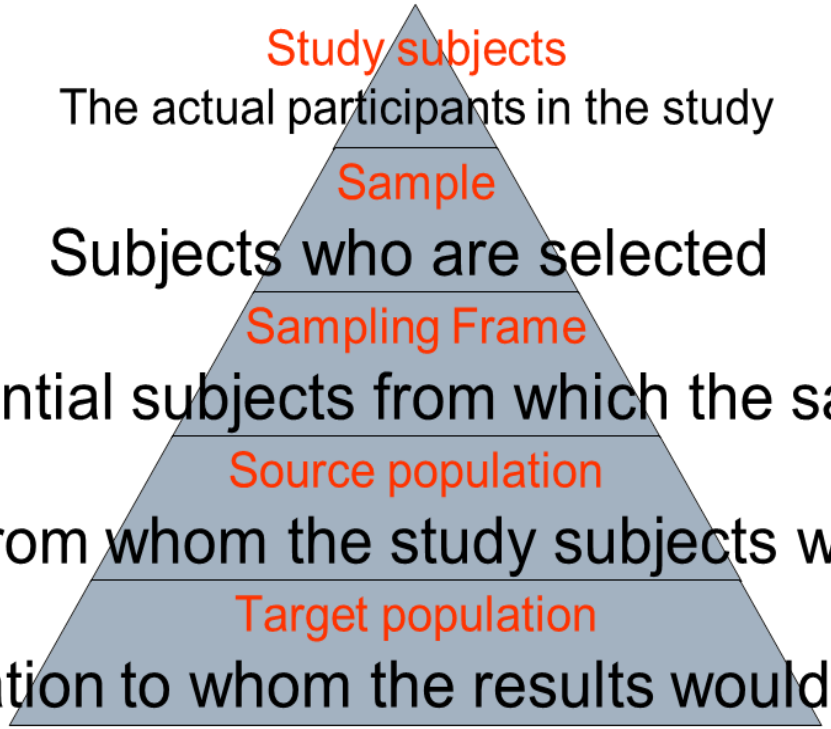
How can you get access to them?

The Sampling Frame

Who is in your study?

The Sample

- **Sampling unit** - the unit of selection in the sampling process
- **Study unit (study subjects)**- the unit on which information is collected or on which observations are made. E.g. Familiar examples are families, towns, litters, branches of a company, individual subjects or schools.


Study subjects
The actual participants in the study
Sample
Subjects who are selected
Sampling Frame
The list of potential subjects from which the sample is drawn
Source population
The Population from whom the study subjects would be obtained
Target population
The population to whom the results would be applied

Target Population:

The population to be studied/ to which the investigator wants to generalize his results

Sampling Unit:

smallest unit from which sample can be selected

Sampling frame

List of all the sampling units from which sample is drawn

Sample size

The number of units in a sample is called the sample size.

Sample Design

A set of rules or procedures that specify how a sample is to be selected. This can either be probability or non-probability

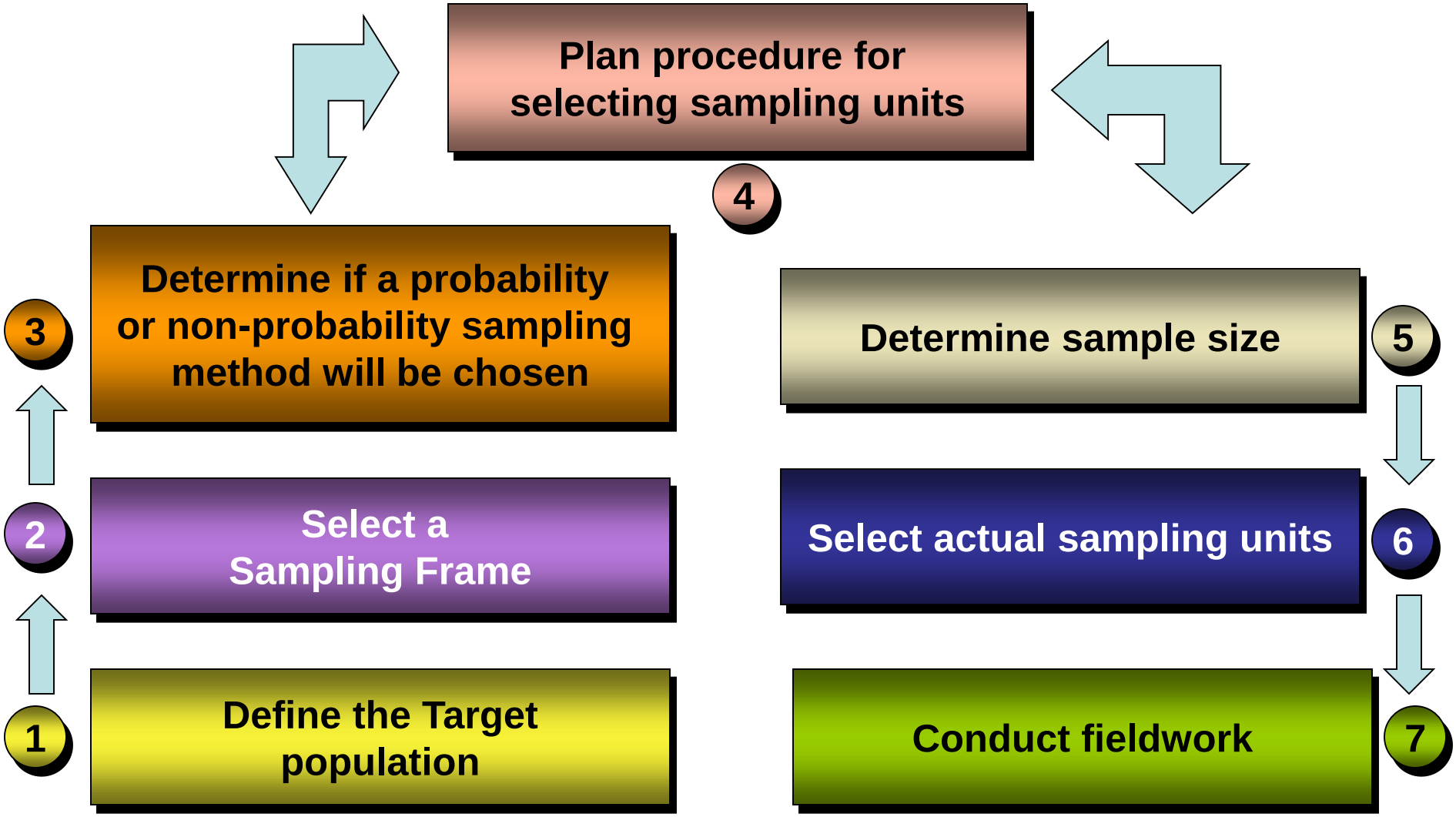
What is sampling ?

- Sampling is simply the process of learning about the population on the basis of sample drawn from it (population). Therefore, in sampling , only part of the population is studied and the conclusions are drawn on that basis for the entire population or universe .
- **A sample is some part of a larger body specially selected to represent the whole**

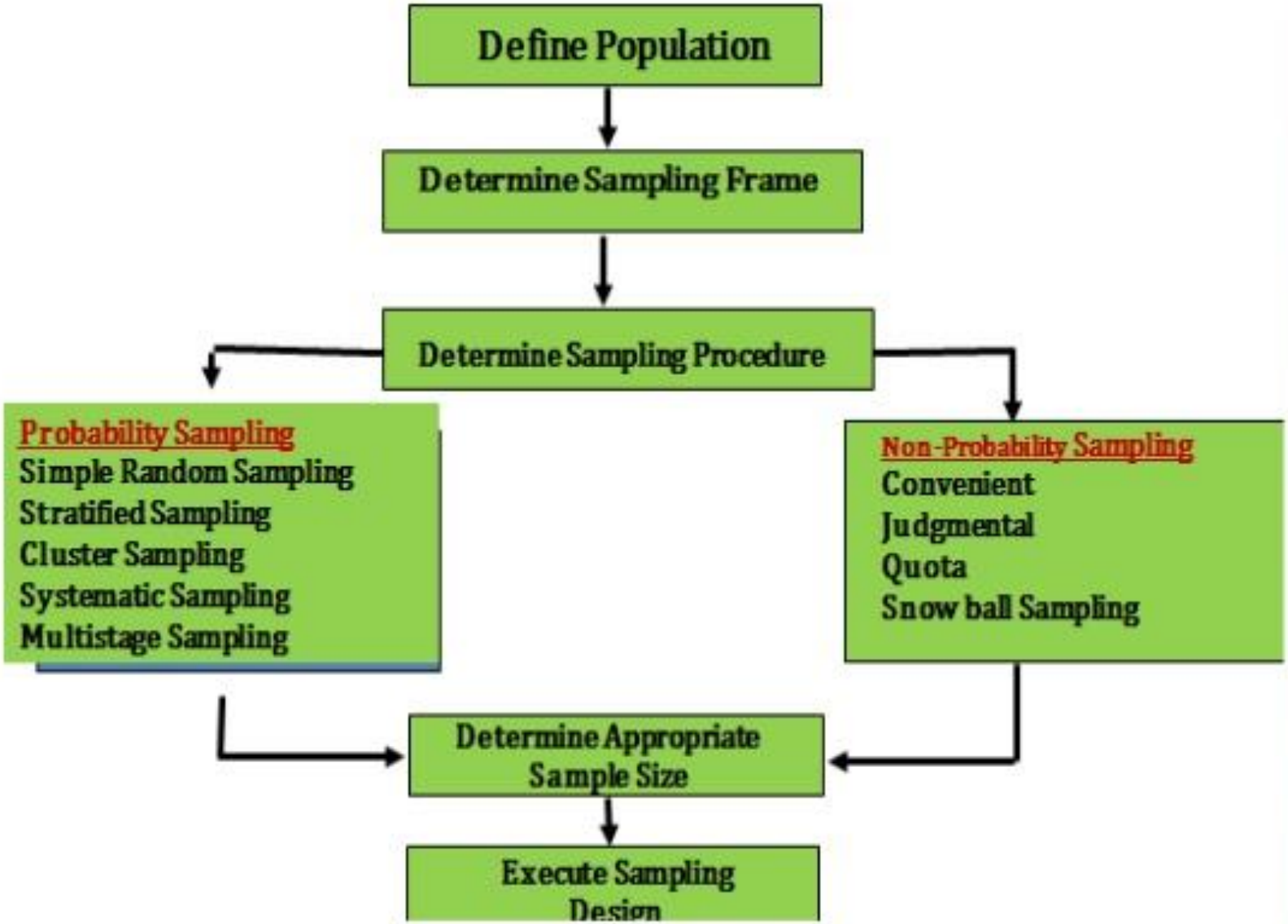
Why Sampling ?

- Less time consuming than a census
- Less cost than a census
- More detailed information
- Less problem and more practical to conduct than a census of the targeted population.

The Sampling Process

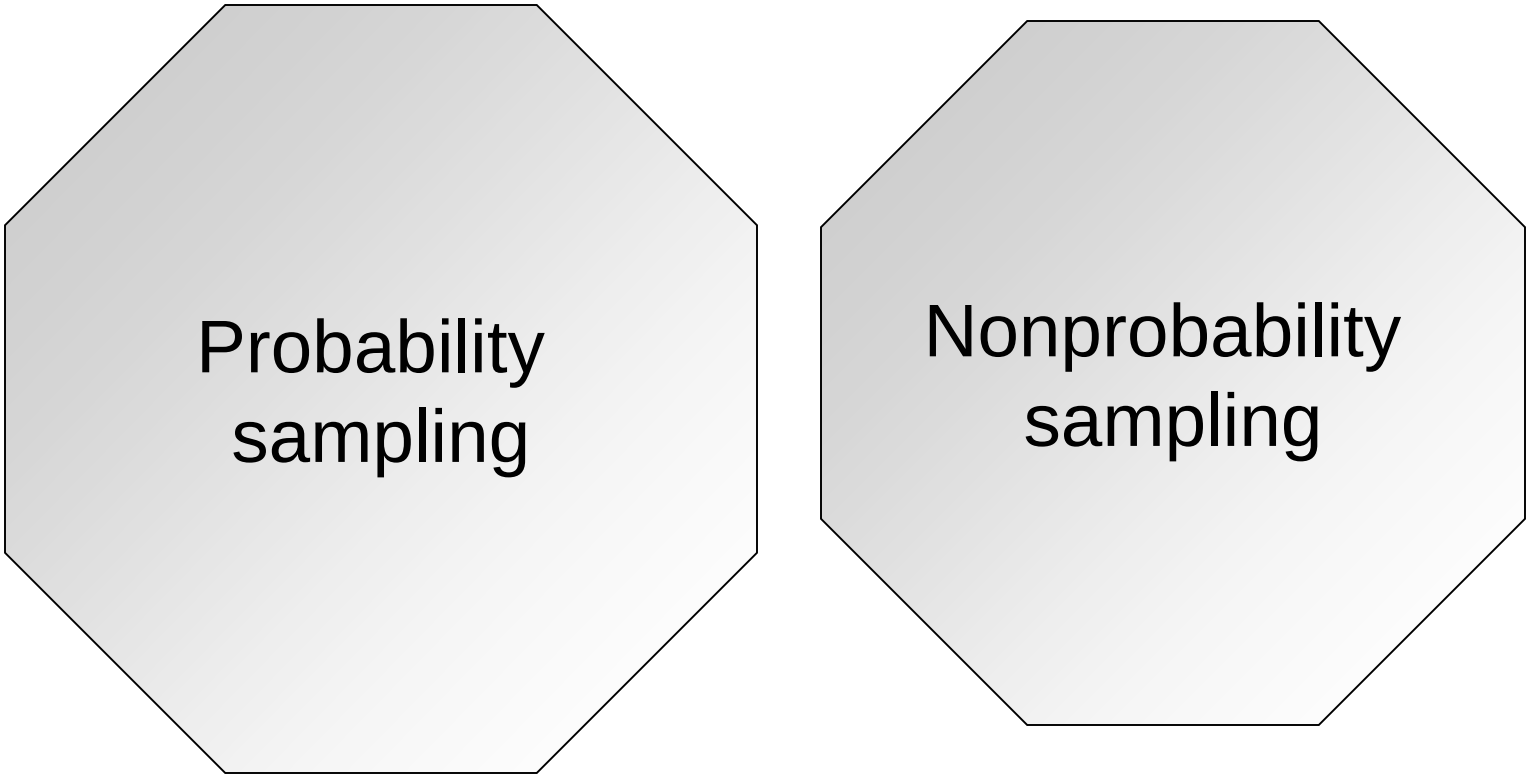


SAMPLING DESIGN PROCESS



Types of Sampling

Mainly sampling is divided into two types



Sampling Design

Probability

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

Nonprobability

- Convenience sampling
- Judgment sampling
- Quota sampling
- Snowball sampling

Probability Sampling

Probability sampling is that sampling in which every items in the universe/population has known chance, of being chosen for the sample . It implies the selection of sample items is independent of person making the study . more to say, the sample is selected on the basis of chance . There is no bias

Non-Probability Sampling

Non- probability sampling methods are those which do not provide every items in the universe/population with a know chance of being included in the sample

SIMPLE RANDOM SAMPLING

- Applicable when population is small, homogeneous & readily available
- All subsets of the frame are given an equal probability. Each element of the frame thus has an equal probability of selection.
- It provides for greatest number of possible samples. This is done by assigning a number to each unit in the sampling frame.

Therefore,

- Samples are selected on the basis of chance
- Personal bias of the investigators does not influence the sample selection
- A table of random number or lottery system is used to determine which units are to be selected.

Lottery method

Under this method,

- all items of the universe are numbered or named on separate slips of paper of identical size and shape.
- These slips are then folded and mixed up in a container.
- A blindfold selection is then made of the number of slips required to constitute the desired sample size.
- The selection items thus depends entirely on chance.

Random number tables

- Random number tables consist of a randomly generated series of digits (0-9).
- To make them easy to read there is typically a space between every 4th digit and between every 10th row.
- When reading from random number tables you can begin anywhere (choose a number at random) but having once started you should continue to read across the line or down a column and NOT jump about.

Random Number Table

13962	70992	65172	28053	02190	83634	66012	70305	66761	88344
43905	46941	72300	11641	43548	30455	07686	31840	03261	89139
00504	48658	38051	59408	16508	82979	92002	63606	41078	86326
61274	57238	47267	35303	29066	02140	60867	39847	50968	96719
43753	21159	16239	50595	62509	61207	86816	29902	23395	72640
83503	51662	21636	68192	84294	38754	84755	34053	94582	29215
36807	71420	35804	44862	23577	79551	42003	58684	09271	68396
19110	55680	18792	41487	16614	83053	00812	16749	45347	88199
82615	86984	93290	87971	60022	35415	20852	02909	99476	45568
05621	26584	36493	63013	68181	57702	49510	75304	38724	15712

Here is an extract from a table of random sampling numbers:

- **3680 2231 8846 5418 0498 5245 7071 2597**
- **If we were doing market research and wanted to sample two houses from a street containing houses numbered 1 to 48 we would read off the digits in pairs**
  36 80 22 31 88 46 54 18 04 98 52 45 70 71 25 97
  and take the first two pairs that were less than 48, which gives house numbers 36 and 22.

- If we wanted to sample two houses from a much longer road with 140 houses in it we would need to read the digits off in groups of three:
  368 <u>022</u> 318 846 541 804 985 245 707 <u>1 25</u> 97
  and the numbers underlined would be the ones to visit: 22 and 125

Systematic Random Sampling

Procedure:

- Number units in population from 1 to N.
- Decide on the n that you want or need.
- K=N/n where k the Interval size, N=Universe size , n=sample size
- Randomly select a number from 1 to k.
- Take every kth unit.

Systematic Random Sampling

This method is used when we have complete list of population under study . This list may be prepared in alphabetical, geographical, numerical or some other order. The first items is selected at random and thereafter at regular intervals.

N = 100

Want n = 20

K=N/n = 5

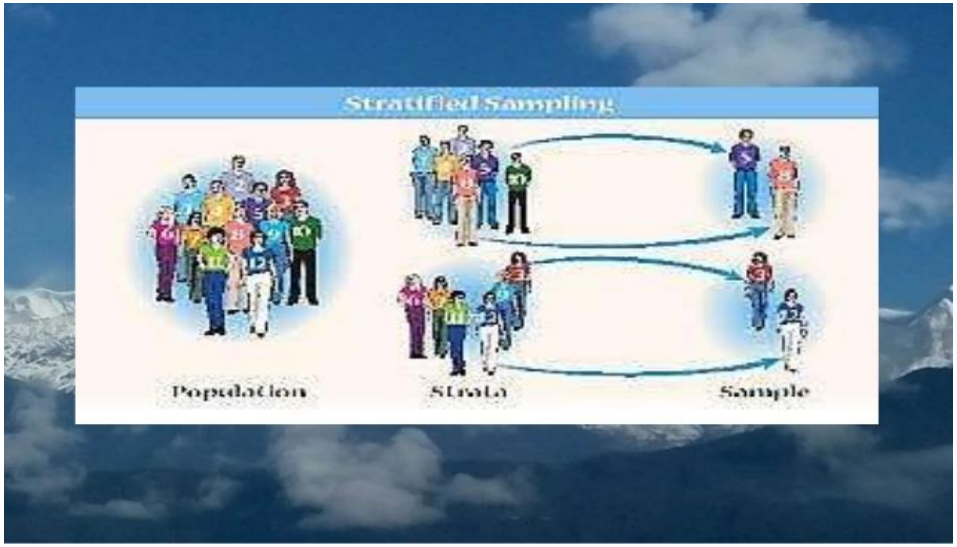
Select a random number from 1-5: chose 4

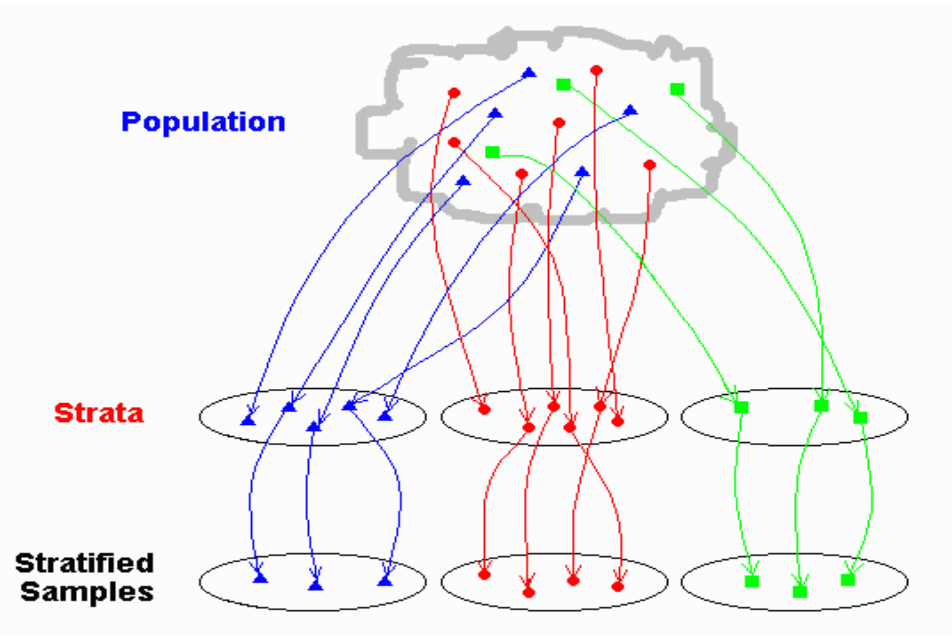
Start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Stratified Sampling

- Stratified sampling techniques are generally used when the population is heterogeneous.
- In stratified sampling, the population is divided into groups called strata.
- A sample is then drawn from within these strata





How to form Strata ?

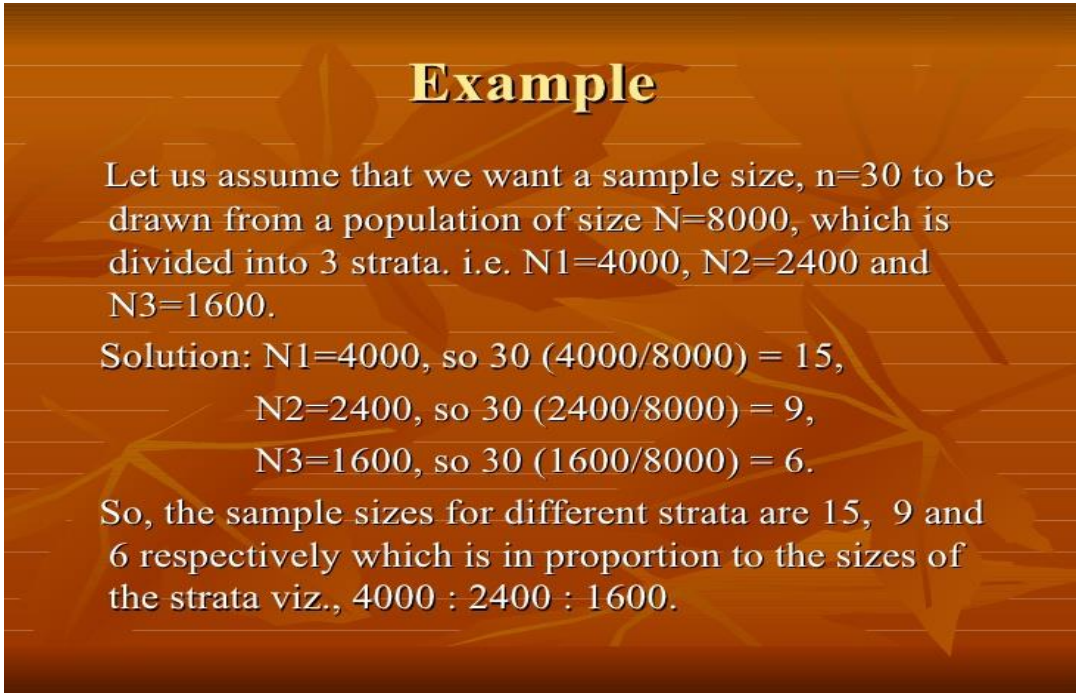
- **Subsets of the listing units in the population**
- **Set of strata must be mutually exclusive and collectively exhaustive**
- **Strata are often based on Variables like**
  - **Income**
  - **Education**
  - **Designation**
  - **age**
  - **sex**
  - **Caste**
  - **Religion**

Sample size within strata

How many sample should be taken from each stratum?

- **Proportional**
- **Disproportional allocation**

In proportional allocation , the number of items drawn from each strata is proportional to the size of the strata.



Cluster (Area) Random Sampling

Procedure:

- Divide population into clusters.
- Randomly sample clusters.
- Measure all units within sampled clusters.

Cluster (Area) Random Sampling

- Advantages: Administratively useful, especially when you have a wide geographic area to cover.

- Examples: Randomly sample from city blocks and measure all homes in selected blocks.

Difference between stratified and cluster sampling

In stratified sampling, the strata are constructed such that they are

- within homogeneous and
- Among heterogeneous

In cluster sampling, the clusters are constructed such that they are

- Within heterogeneous and
- Among homogeneous

- All strata are represented in the sample; but only a subset of clusters are in the sample

MULTI-STAGE SAMPLING

- The procedure of first selecting large sized units and then choosing a specified number of sub-units from the selected large units is known as sub-sampling.
- The large units are called ‘first stage units’ and the sub-units the ‘second stage units’.
- The procedure can be easily generalised to three stage or multistage samples.

Convenience Sampling

- Sometimes known as **grab** or **opportunity sampling or accidental or haphazard sampling.**
- Selection of whichever individuals are easiest to reach.
- It is done at the “convenience” of the researcher.
- **Researcher tend to make the selection at familiar locations and to choose respondents who are like themselves.**
  - **Error occurs 1) in the form of members of the population who are infrequent or nonusers of that location and 2) who are not typical in the population**

Purposive/Judgmental Sampling

- ***Judgment samples:*** **samples that require a judgment or an "educated guess" on the part of the interviewer as to who should represent the population.**

Also, "judges" (informed individuals) may be asked to suggest who should be in the sample.

  - **Subjectivity enters in here, and certain members of the population will have a smaller or no chance of selection compared to others**

QUOTA SAMPLING

- The population is first segmented into mutually exclusive sub-groups, just as in stratified sampling.
- Then judgment used to select subjects or units from each segment based on a specified proportion.
- For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.
- It is this second step which makes the technique one of non-probability sampling.
- In quota sampling the selection of the sample is non-random.
- For example interviewers might be tempted to interview those who look most helpful. The problem is that these samples may be biased because not everyone gets a chance of selection. This random element is its greatest weakness and quota versus probability has been a matter of controversy for many years

Snowball Sampling

- **Used in studies involving respondents who are rare to find.**
- **To start with, the researcher compiles a short list of sample units from various sources.**
- **Each of these respondents are contacted to provide names of other probable respondents.**

Potential Sources of Error in Research Design

- **Total Error**
  - **Random Sampling Error**
  - **Non-sampling Error**
    - **Response Error**
      - **Researcher Error**
        - Surrogate Information Error
        - Measurement Error
        - Population Definition Error
        - Sampling Frame Error
        - Data Analysis Error
      - **Interviewer Error**
        - Respondent Selection Error
        - Questioning Error
        - Recording Error
        - Cheating Error
      - **Respondent Error**
        - Inability Error
        - Unwillingness Error
    - **Non-response Error**

What size sample do I need?”

- The size of the universe/population
- The resources available
- The degree of accuracy desired
- Homogeneity or Heterogeneity of the universe
- Nature of study
- Nature of respondents
- Type of analysis to be employed
- The level of precision needed
- Sampling technique used

Calculating Sample Size.

There are different procedures that could be used for calculating sample size:

- Use of formulae
- Ready made table
- Computer software

Sample size determination in quantitative study

Several criteria will need to be specified to determine the appropriate sample size:

- Level of precision/ **Sampling Error**,
- Level of confidence or risk,
- **Degree of Variability**

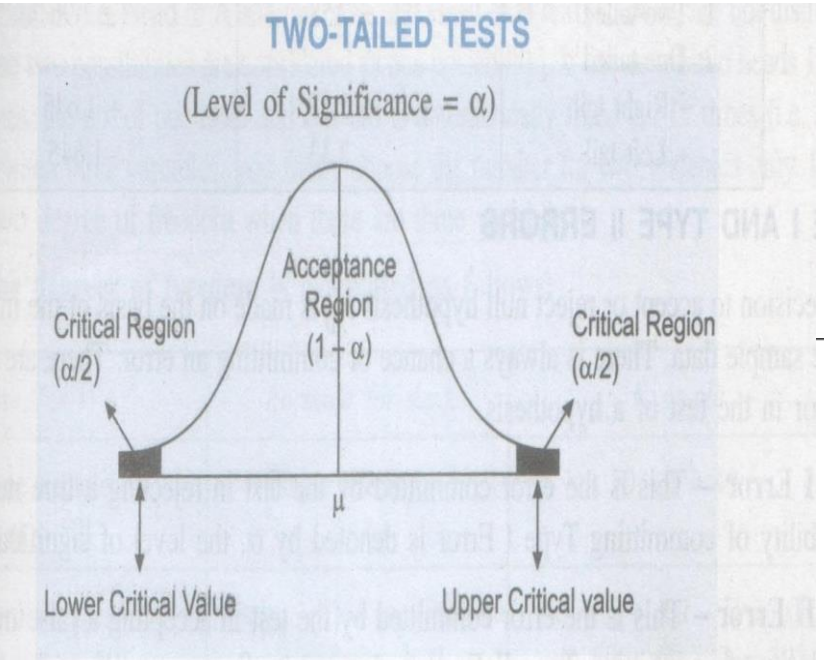
Level of precision

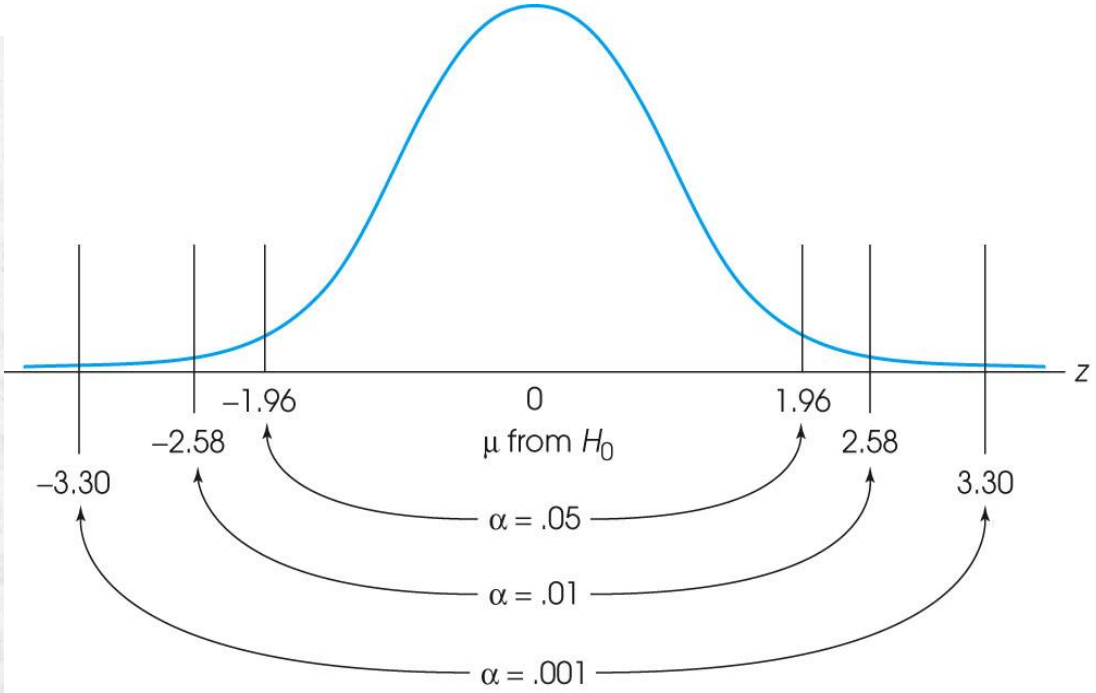
- Sample size is to be determined according to some pre assigned "degree of precision"
- The 'degree of precision' is the margin of permissible error between the estimated value and the population value.
- In other words, it is the measure of how close an estimate is to the actual characteristic in the population.
- **The Level of Precision-***sometimes called* ***sampling error/*** 'confidence interval'
- The difference between the sample statistic and the related population parameter is called the sampling error

- Range in which the true value of the population is estimated to be.
  - This range is often expressed in percentage points (e.g., ±5 percent).
  - If the sampling error or margin of error is ±5%, and 70% unit in the sample attribute some criteria, then it can be concluded that 65% to 75% of units in the population have attributed that criteria
- High level of precision requires larger sample sizes and higher cost to achieve those samples.

The Confidence Level / Risk Level

- How confident do you want to be that the actual mean falls within your confidence interval?
- The most common confidence intervals are **90%** confident, **95%** confident, and 99% confident
  - E.g. a 95% confidence level is selected, 95 out of 100 samples will have the true population value within the range of precision





Degree of Variability

- refers to the distribution of attributes in the population.
- The more heterogeneous a population, the larger the sample size required to obtain a given level of precision.
- The less variable (more homogeneous) a population, the smaller the sample size.
- You should note that a 50/50 split on a specific attribute or response indicates maximum variability in the population, whereas a 90/10 split means that 90 per cent of the population share an attribute, so the sample is less variable.
- If you don't know what level of variability to expect, then assume that it is 50 per cent (.5)
- This may mean that you use a larger sample size than was really needed, but that is better than using a sample size that is too small, and then having no confidence in the results.

Cochran's formula for calculating sample size when the population is infinite:

Cochran (1977) developed a formula to calculate a representative sample for proportions as

$$n_o = \frac{Z^2 pq}{e^2}$$

Where n_0 is the sample size,

- Z^2 is the abscissa of the normal curve that cuts off an area α at the tails; (1 – α) equals the desired confidence level, e.g., 95%);
- e is the desired level of precision/Margin of error
- p is the estimated proportion of an attribute that is present in the population (degree of Variability), and q is 1-p.

Note:

$$n_o = \frac{Z^2 pq}{e^2}$$

- p=proportion in the target population estimated to have a particular characteristics. If there is no reasonable estimate, use 50%(i.e 0.5)
- q=1-p(proportion in the target population not having the particular characteristics)
- Z= The value for Z is found in statistical tables which contain the area under the normal curve. e.g. Z = 1.96 for 95 % level of confidence

- e= degree of accuracy required, usually set at 0.05 level(occasionally at 2.0)

Example

Suppose we want to calculate a sample size of a large population whose degree of variability is not known. Assuming the maximum variability, which is equal to 50% (p =0.5) and taking 95% confidence level with ±5% precision, the calculation for required sample size will be as follows--

Sample size is

$p = 0.5$ and hence q =1-0.5 = 0.5; $e = 0.05$; z =1.96

$$n_0 = \frac{Z^2 pq}{e^2}$$ (1st equation)

$$\text{So, } n_0 = \frac{(1.96)^2 (0.5)(0.5)}{(0.05)^2} = 384.16 = 384$$

Again, taking 99% confidence level with ±5% precision, the calculation for required sample size will be as follows--

$p = 0.5$ and hence q=1-0.5 = 0.5; e= 0.05; z=2.58

$$\text{So, } n_0 = \frac{(2.58)^2 (0.5)(0.5)}{(0.05)^2} = 665.64 = 666$$

Sample size calculated for different confidence level and precision

Confidence level	Sample size (n_0)		
	e =.03	e =.05	e = .1
95%	1067	384	96
99%	1849	666	166

Cochran's formula for calculating sample size when the population is finite

- Cochran pointed out that if the population is finite, then the sample size can be reduced slightly.
- This is due to the fact that a very large population provides proportionally more information than that of a smaller population.
- He proposed a correction formula to calculate the final sample size in this case which is given below

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

- Here, n_0 is the sample size derived from equation (1) and *N* is the population size

- Now, suppose we want to calculate the sample size for the population of our study where, population size is *N* =13191.
- According to the formula (1), the sample size will be 666 at 99% confidence level with margin of error equal to (0.05).
- If $\frac{n_0}{N}$ is negligible then n_0 is a satisfactory approximation to the sample size.

- But in this case, the sample size (666) exceeds 5% of the population size (13191).
- So, we need to use the correction formula to calculate the final sample size
- Here, *N* = 13191, 0 *n* = 666 (using formua 1)

$$n = \frac{666}{1 + \frac{(666-1)}{13191}} = 634.03 = 634$$

- But, if the sample size is calculated at 95% confidence level with margin of error equal to (0.05), the sample size become 384 which does not need correction formula. So, in this case the representative sample size for our study is 384

Yamane's formula for calculating sample size

- Yamane suggested another simplified formula for calculation of sample size from a population which is an alternative to Cochran's formula
- According to him, for a 95% confidence level and $p = 0.5$, size of the sample should be

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

where, N is the population size and e is the level of precision .

- Let this formula be used for our population, in which *N* =13191 with ±5% precision.
- Assuming 95% confidence level and *p* =0.5, we get the sample size as

$$n = \frac{13191}{1 + 13191(.05)^2} = 388$$

Sample sizes calculated by Yamane's formula

Sl. no. of schools	Population size,N	Sample size, *n* for 95% confidence level:		
		±5%	±7%	±10%
1	450	212	136	82
2	582	229	150	85
3	693	254	158	87
4	799	266	163	89
5	806	267	163	89
6	845	272	164	89
7	858	273	165	90
8	892	276	166	90
9	909	278	167	90
10	922	279	167	90
11	9 85	285	169	91
12	1009	287	170	91
13	1058	290	171	91
14	1073	292	171	91
15	1115	294	173	92
16	1167	299	174	92
17	1184	299	174	92
18	1256	303	176	93
19	1298	305	176	93
20	1322	307	177	93
21	1584	319	181	94
22	1908	330	184	95

Sample sizes calculated by Cochran's formula

Sl.no. of schools	Population size,N	Sample size, *n* at 95% confidence level:			Sample size, *n* at 99% confidence level:		
		±5%	±7%	±10%	±5%	±7%	±10%
1	450	208	137	79	269	194	121
2	582	231	146	83	311	215	130
3	693	248	153	84	340	228	134
4	799	259	158	86	364	239	137
5	806	259	158	86	364	239	137
6	845	265	159	86	372	243	138
7	858	265	161	86	374	243	139
8	892	269	161	86	381	246	141
9	909	270	162	87	385	248	141
10	922	270	162	87	387	248	141
11	9 85	276	163	87	396	253	142
12	1009	278	165	88	398	253	143
13	1058	282	166	88	409	256	143
14	1073	282	166	88	411	256	144
15	1115	286	168	88	416	262	144
16	1167	289	168	89	424	264	146
17	1184	291	169	89	427	264	146
18	1256	295	169	89	435	268	147
19	1298	295	170	90	441	270	147
20	1322	298	170	90	444	270	148
21	1584	310	175	91	469	281	151
22	1908	320	178	91	493	288	152

Note

- The sample size formulas provide the number of responses that need to be obtained. Many researchers commonly add 10 % to the sample size to compensate for persons that the researcher is unable to contact.
- The sample size also is often increased by 30 % to compensate for non-response (e.g self administered questionnaires).

Use Of Readymade Table For Sample Size Calculation

- How large a sample of patients should be followed up if an investigator wishes to estimate the incidence rate of a disease to within 10% of it's true value with 95% confidence?
- The table show that for e=0.10 & confidence level of 95%, a sample size of 385 would be needed.
- This table can be used to calculate the sample size making the desired changes in the relative precision & confidence level .e.g if the level of confidence is reduce to 90%, then the sample size would be 271.
- Such table that give ready made sample sizes are available for different designs & situation

Table 1. Sample size for ±3%, ±5%, ±7% and ±10% Precision Levels Where Confidence Level is 95% and P=.5.

Size of Population	Sample Size (n) for Precision (e) of:			
	±3%	±5%	±7%	±10%
500	a	222	145	83
600	a	240	152	86
700	a	255	158	88
800	a	267	163	89
900	a	277	166	90
1,000	a	286	169	91
2,000	714	333	185	95
3,000	811	353	191	97
4,000	870	364	194	98
5,000	909	370	196	98
6,000	938	375	197	98
7,000	959	378	198	99
8,000	976	381	199	99
9,000	989	383	200	99
10,000	1,000	385	200	99
15,000	1,034	390	201	99
20,000	1,053	392	204	100
25,000	1,064	394	204	100
50,000	1,087	397	204	100
100,000	1,099	398	204	100
>100,000	1,111	400	204	100

a = Assumption of normal population is poor (Yamane, 1967). The entire population should be sampled.

Table 2. Sample size for ±5%, ±7% and ±10% Precision Levels Where Confidence Level is 95% and P=.5.

Size of Population	Sample Size (n) for Precision (e) of:		
	±5%	±7%	±10%
100	81	67	51
125	96	78	56
150	110	86	61
175	122	94	64
200	134	101	67
225	144	107	70
250	154	112	72
275	163	117	74
300	172	121	76
325	180	125	77
350	187	129	78
375	194	132	80
400	201	135	81
425	207	138	82
450	212	140	82

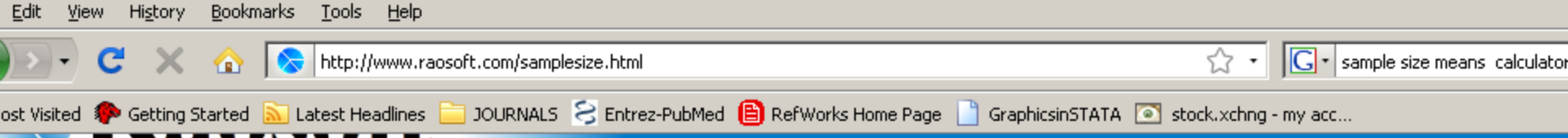
Table 3: Estimating an incidence rate with specified relative precision [Formula: $n = (Z_{1-\alpha/2} / \varepsilon)^2$]

Relative precision (ε)	Confidence level		
	99%	95%	90%
0.01	66358	38417	27061
0.02	16590	9605	6766
0.03	7374	4269	3007
0.04	4148	2402	1692
0.05	2655	1537	1083
0.06	1844	1068	752
0.07	1355	785	553
0.08	1037	601	423
0.09	820	475	335
0.10	664	385	271
0.12	461	267	188
0.14	339	197	139
0.16	260	151	106
0.18	205	119	84
0.20	166	97	68
0.22	138	80	56
0.24	116	67	47
0.26	99	57	41
0.28	85	50	35
0.30	74	43	31
0.32	65	38	27
0.34	58	34	24
0.36	52	30	21
0.38	46	27	19
0.40	42	25	17
0.42	38	22	16
0.44	35	20	14
0.46	32	19	13
0.48	29	17	12
0.50	27	16	11

USE OF COMPUTER SOFTWARE FOR SAMPLE SIZE CALCULATION & POWER ANALYSIS

The following software can be used for calculating sample size & power;

- Epi-info
- nQuerry
- Power & precision
- Sample
- STATA
- SPSS



t margin of error can you accept? is a common choice	5 %	The margin of error is the amount of error that you can tolerate. If 90% of respondents answer *yes*, while 10% answer *no*, you may be able to tolerate a larger amount of error than if the respondents a split 50-50 or 45-55. Lower margin of error requires a larger sample size.
t confidence level do you need? ical choices are 90%, 95%, or 99%	95 %	The confidence level is the amount of uncertainty you can tolerate. Suppose that you have 20 yes-n questions in your survey. With a confidence level of 95%, you would expect that for one of the questions (1 in 20), the percentage of people who answer *yes* would be more than the margin of err away from the true answer. The true answer is the percentage you would get if you exhaustively interviewed everyone. Higher confidence level requires a larger sample size.
t is the population size? ou don't know, use 20000	20000	How many people are there to choose your random sample from? The sample size doesn't change much for populations larger than 20,000.
at is the response distribution? ave this as 50%	50 %	For each question, what do you expect the results will be? If the sample is skewed highly one way or the other,the population probably is, too. If you don't know, use 50%, which gives the largest sample size.
recommended sample size is	**377**	This is the minimum recommended size of your survey. If you create a sample of this many people and get responses from everyone, you're more likely to get a correct answer than you would from a large sample where only a small percentage of the sample responds to your survey.

ernate scenarios

With a sample size of	100	200	300	With a confidence level of	90	95	99
Your margin of error would be	9.78%	6.89%	5.62%	Your sample size would need to be	267	377	643

Questions/Clarification