Learning Outcome

By the end of this lecture, you will be able to:

- Describe different sampling methods

Introduction

Let's consider Ahmad, a Quality Engineer in a manufacturing company. Ahmad is willing to examine overall production quality of goods in his company. The company has four production lines. Will collecting data from one production line help examine overall production quality?

Ahmad, Quality Engineer

Production line

Certainly not! Different types of sampling techniques should be used for different needs.

In this lecture, we will learn about the different sampling methods.

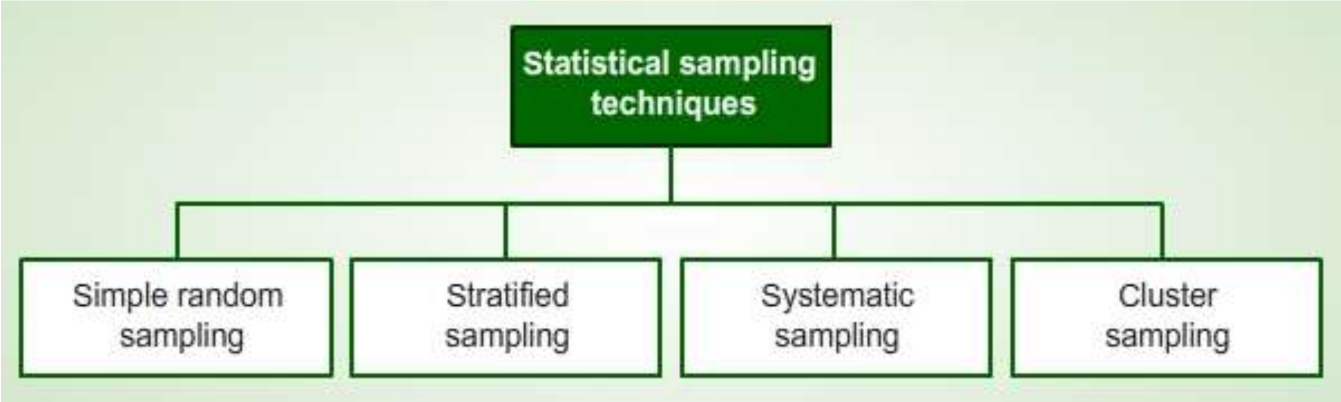
Sampling Techniques

There are two categories of sampling techniques:

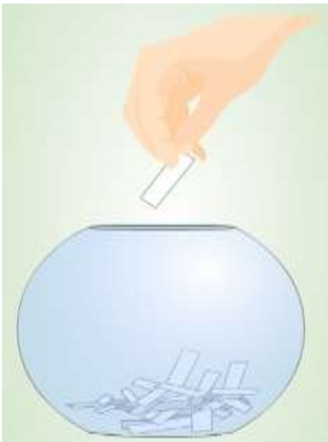
Non statistical sampling technique	Non statistical sampling techniques or non-probability sample are used for judgment or convenience.
Statistical sampling technique	Statistical sampling techniques or probability sample are chosen based on known or calculable probabilities.

Statistical Sampling Techniques

Some statistical sampling techniques are:



Simple Random Sampling



Random sampling

A sample is called a simple random sample if each unit of the population has an equal chance of being selected for the sample. So that each possible sample has the same probability of being chosen, also every item in the sample has the same chance to being selected.

Selection may be made with replacement or without replacement (small populations) and can be obtained from a table of random numbers or computer random number generators.

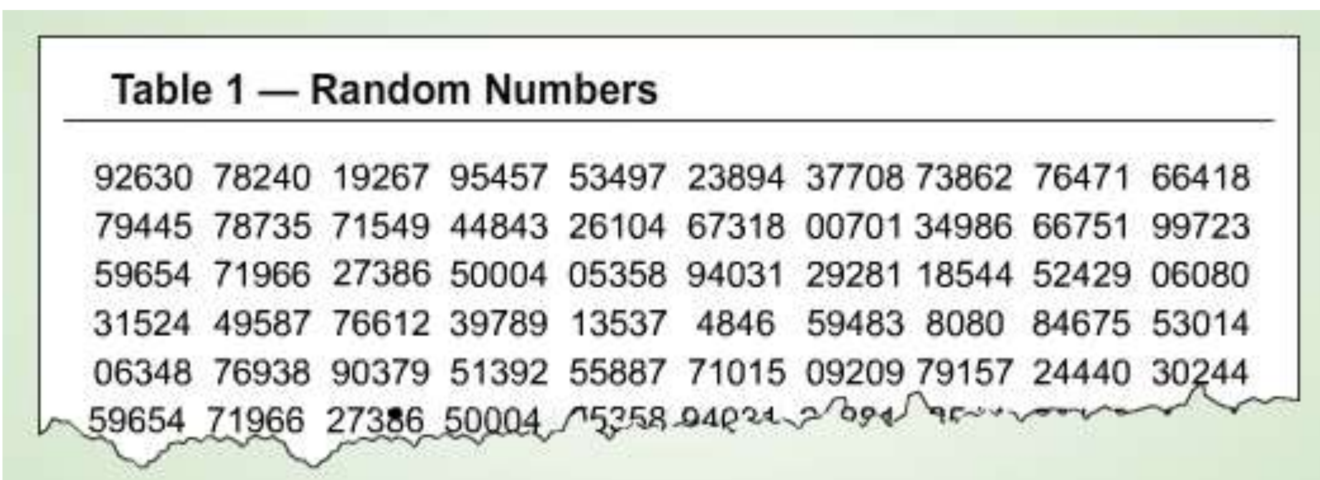


Table 1 — Random Numbers

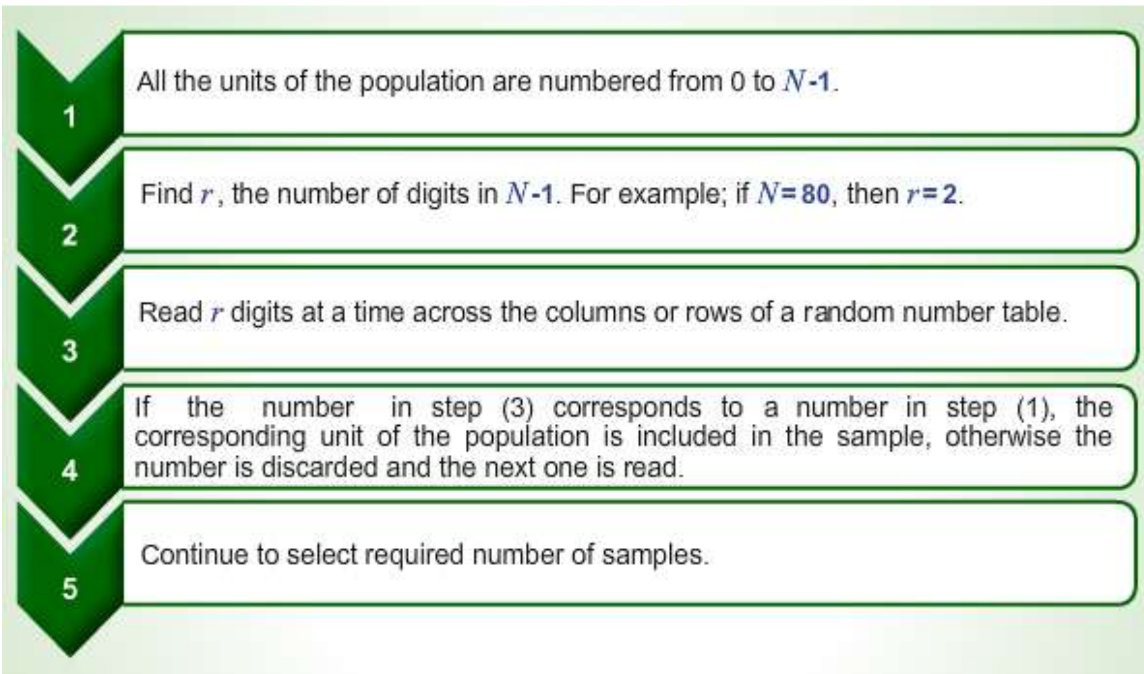
92630	78240	19267	95457	53497	23894	37708	73862	76471	66418
79445	78735	71549	44843	26104	67318	00701	34986	66751	99723
59654	71966	27386	50004	05358	94031	29281	18544	52429	06080
31524	49587	76612	39789	13537	4846	59483	8080	84675	53014
06348	76938	90379	51392	55887	71015	09209	79157	24440	30244
59654	71966	27386	50004	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Simple random sampling

Drawing Simple Random Samples using a Table of Random Numbers

An easy way to select a SRS is to use a random number table, which is a table of digits 0,1,…,9. Here, each digit has an equal chance of being selected at each draw.

To use this table in drawing a random sample of size n from a population of size N, these are the steps performed:



If the same unit in the population is selected more than once in the above process of selection, then the resulting sample is called a **SRS with replacement**; otherwise it is called a **SRS without replacement**. The observations in the sample are the enumeration or readings of the units selected.

Drawing Simple Random Samples >Example

The table shows branch codes for a company in a country given according the provinces in which the branches are located.

Branch codes for a company in a country given according the provinces

West				East			South		North
AB$^{(00)}$	CD$^{(01)}$	EF	GH	IJ	KL	MN	OP	QR	ST$^{(09)}$
UV$^{(10)}$	WX$^{(11)}$	YZ	ZA	BY	XC	WD	VE	UF	GT$^{(19)}$
SH	RI	QJ	PK	OL	NM	AC	EG	IK	MO
QS	UW	YB	DF	HJ	LN	PR	TV	XZ	NA
OB	PC	QD	RE	SF	TG	UH	VI	WJ	XK
YL	ZM	AV	CU	ET	GS	IR	KQ	MP	OO
QN	SM	UL	WK	YJ	BI	HD	FG	HF	JE
LD	NC	PB	RA	TA	VB	CX	ZD	AE	BF
CG$^{(80)}$	DH	EI	FJ	GK	HL	IM	JN	KO	LP$^{(89)}$
MQ$^{(90)}$	NR	OS	PT	QU	RV	SW	TX	UY	VZ$^{(99)}$

Starting from the upper left corner, moving column by column, and using the part of the random number table given below.

Part of a random number table

8571	7683	5118	7669	6126	3663	3059	7807	9219	4383	9021	7013	0233	3348	4077
0864	5055	8631	5770	0505	0386	9792	1690	4874	3084	0228	8539	9375	5046	8635
4753	1992	8182	2658	2914	4005	1577	1714	7862	7009	0252	3070	1563	3008	3716
1267	1063	4415	8496	6779	1563	7833	5351	2278	0674	1252	6813	4016	3961	6890
9497	0105	5626	0529	0602	4573	1499	7772	7759	9405	9502	3408	6931	7946	4655
6823	7365	6140	0357	7069	7715	9083	6180	1131	7059	9808	9803	7883	5943	6649
6532	4048	3044	8035	1045	8349	5422	0315	7470	7679	1726	1390	4997	5632	9033
8184	8336	5684	5846	7056	2847	4715	2869	2576	5373	8175	0384	5348	8232	8186
5605	0939	9380	1647	7307	5893	7569	7092	4437	2722	7807	5908	5425	9679	2348
4926	1561	7299	2195	5374	3664	8269	5241	4436	5265	7571	8299	6006	2142	2273
0933	6131	2406	0715	5069	1663	8015	9120	0667	4884	8601	3370	3449	7158	8950
7413	9526	9670	3075	8321	8295	6327	5475	5650	9061	7687	3849	2207	6910	4166

Solution:

To draw a simple random sample of size 10 from the branches in the four provinces (without replacement), we proceed as follows:

1. The size of the population is 100, so label the units in the population from 00 to 99.

2. Find r, the number of digits in N-1. For example, if N-1=99, then r=2.

3. Read 2 digits at a time across the columns or rows of a random number table.

   Suppose we read the first two digits of the first two columns of the above random number table to get the following numbers

   85 71 76 83 51 18 76 69 61 26 36

4. Since the random digit 85 corresponds to a unit, we select unit 85 of the population in the sample. The seventh random number 76 is discarded for *SRS* without replacement as it appeared before.

5. Continue until 10 units have been selected as n is 10. Thus we have the sample units:

   85 71 76 83 51 18 69 61 26 36

   So the branches selected for sample observations are:

   HL NC CX FJ ZM UF JE SM AC PR

   A *SRS* with replacement in the above example would be:

   HL NC CX FJ ZM UF CX JE SM AC

Simple Random Sampling >Advantages and Disadvantages

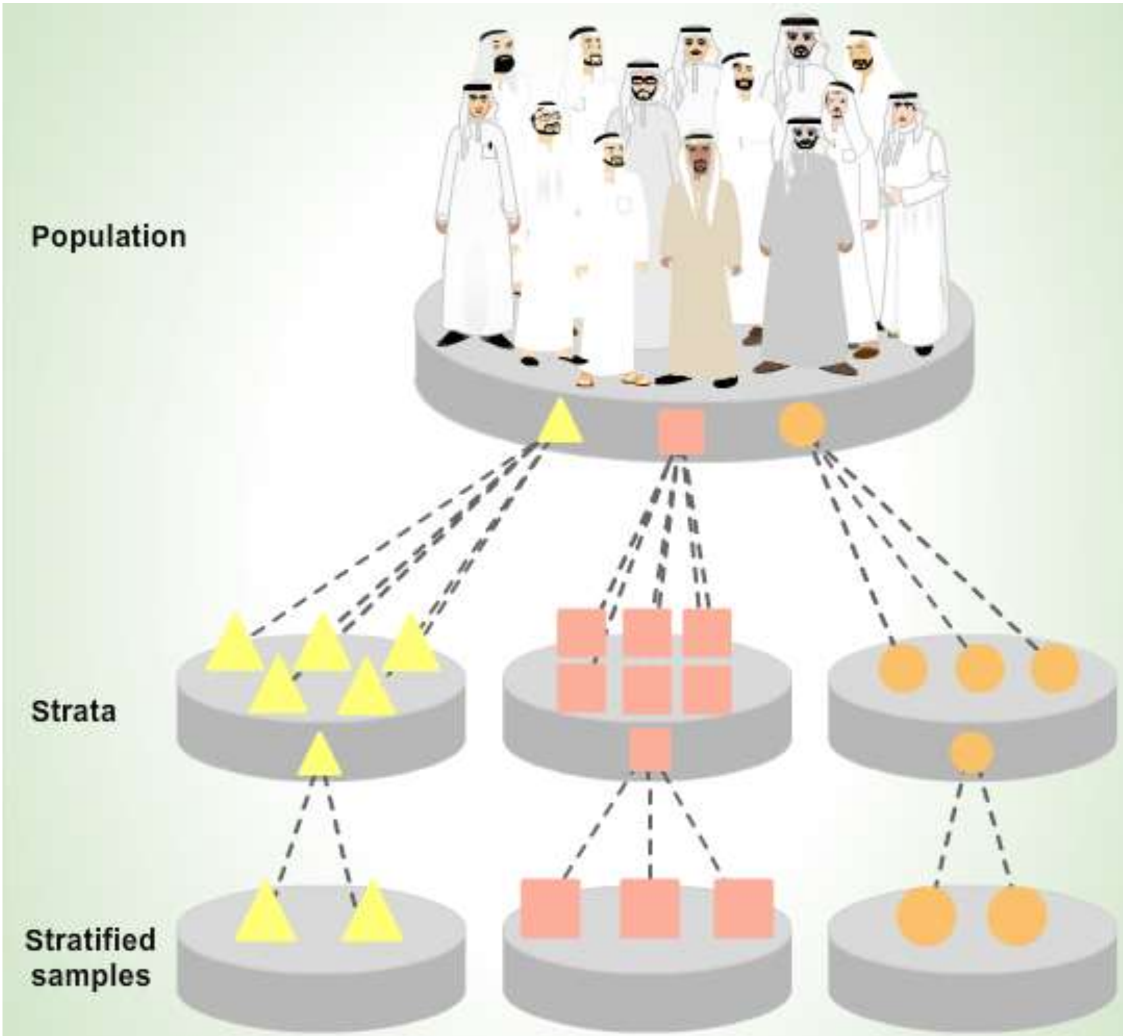
Given here are some of the advantages and disadvantages of simple random samples.

Advantages	SRS is: • Free of classification error • Requires minimum advance knowledge of the population • Best in situations where the population is fairly homogeneous and not much information is available about the population
Disadvantages	If the population of interest can be divided into subgroups, then it is better to draw a stratified or cluster sample to have a representative sample.

Stratified Sampling

Stratification is the process of grouping members of the population into relatively homogeneous subgroups before sampling. The population divided into subgroups (called strata) according to some common characteristic such that every element in the population must be assigned to only one stratum. The strata should be mutually exclusive and collectively exhaustive.
To develop strata, we look at the characteristic of interest for which items are quite homogeneous.

A *SRS* selected from each stratum, then the sample from all strata are combined into one sample. The main objective is to increase precision by keeping the sample size as small as possible.



Stratified sampling

Given here are some of the advantages and disadvantages of stratified sampling.

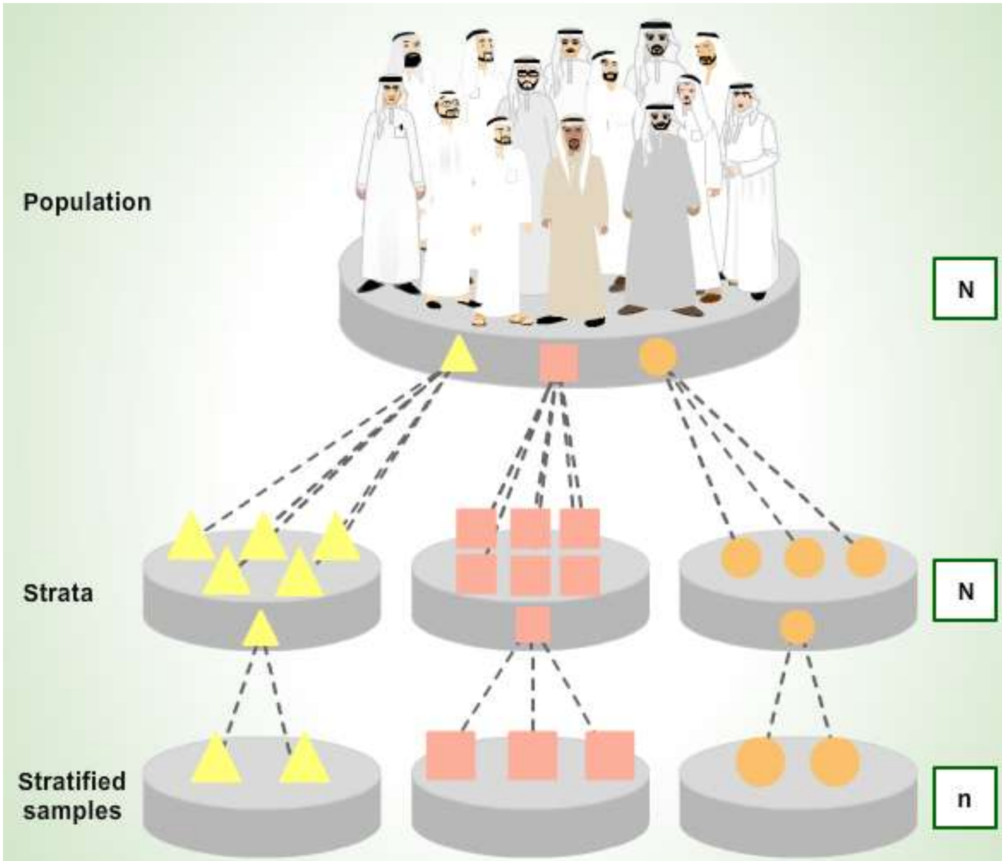
Advantages	Stratified sampling: • Focuses on important subpopulations and ignores irrelevant ones • Improves the accuracy and efficiency of estimation
Disadvantages	Stratified sampling: • Can be complex to use due to difficulty in selecting relevant stratification variables • Cannot be used when there are no homogeneous subgroups • Can be expensive • Requires accurate information about the population, or introduces bias • Looks random within specific sub headings

Drawing Stratified Random Samples

Let us consider a population of size N divided into k subpopulations of sizes N_1, $N_2 \ldots N_k$.

These subpopulations (strata) are disjoint and verify that $N_1 + N_2 + \ldots + N_k = N$.

Each of the subpopulations is called stratum. If we want to have a sample of size n elements from the initial population, we select a simple random sample of size n_i from each stratum so that $n_1 + n_2 + \ldots + n_k = n$.



Stratified sampling

Strategies of Drawing Stratified Random Samples

There are several strategies to draw stratified random samples. Here are two commonly used strategies.

Proportional allocation	Proportional allocation uses a sampling fraction ($n_i = n*N_i/N$) in each of the strata that is proportional to that of the total population. If the population consists of 60% in the male stratum and 40% in the female stratum, then the relative size of the two samples (male and female) should reflect this proportion. In general, the size of the sample in each stratum is taken in proportion to the size of the stratum. This is called proportional allocation.
Optimum allocation	Optimum allocation (or Disproportional allocation) - Each stratum is proportional to the standard deviation of the distribution of the variable. Larger samples are taken in the strata with the greatest variability to generate the least possible sampling variance.

Example

Let us consider the same example of simple random samples. Suppose, we need to draw a stratified random sample of size 10 from the branches in the four provinces (without replacement).

The table shows the number of branches in each province:

Province	Number of branches
West	40
East	30
South	20
North	10

The total is 100 branches and we are asked to take a sample of 10 from the 100 branches in the four provinces (without replacement).The first step is to find the total number of branches (100) and calculate the percentage in each group or stratum.

Province	Number of branches	Percentage of each group or stratum
West	40	$= \left({}^{40}/_{100}\right) \times 100\% = 0.4 \times 100\% = 40\%$
East	30	$= \left({}^{30}/_{100}\right) \times 100\% = 0.3 \times 100\% = 30\%$
South	20	$= \left({}^{20}/_{100}\right) \times 100\% = 0.2 \times 100\% = 20\%$
North	10	$= \left({}^{10}/_{100}\right) \times 100\% = 0.1 \times 100\% = 10\%$

Now, we can calculate number samples from each province for total sample size of 10:

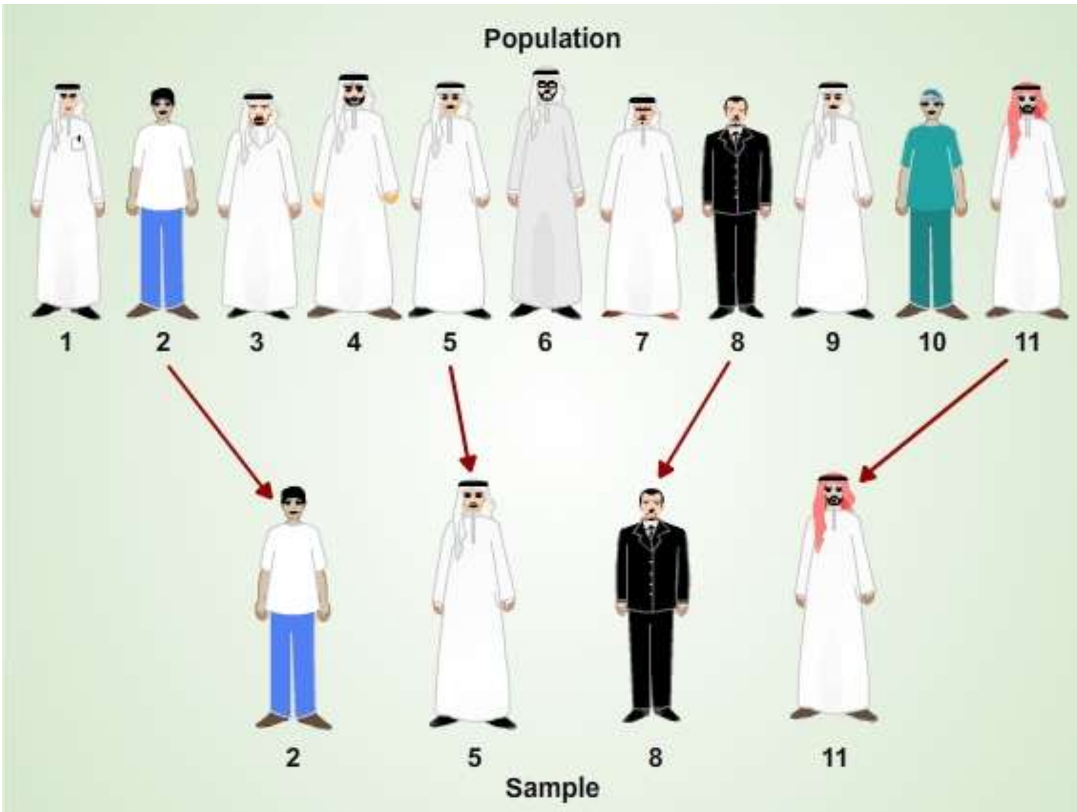
Province	Number of branches	Percentage of each group or stratum	Number of samples to be collected from each province
West	40	$= \left({}^{40}/_{100}\right) \times 100\% = 0.4 \times 100\% = 40\%$	40% of 10 = 4
East	30	$= \left({}^{30}/_{100}\right) \times 100\% = 0.3 \times 100\% = 30\%$	30% of 10 = 3
South	20	$= \left({}^{20}/_{100}\right) \times 100\% = 0.2 \times 100\% = 20\%$	20% of 10 = 2
North	10	$= \left({}^{10}/_{100}\right) \times 100\% = 0.1 \times 100\% = 10\%$	10% of 10 = 1

Select *SRS* from each stratum, then the final sample is the sum of all sub–samples that is 10.

Systematic Sampling

Systematic sampling is a procedure in which each element in the population has a known and equal probability of selection. This makes systematic sampling functionally similar to simple random sampling. It is however, much more efficient and much less expensive to do.

The researcher must ensure that the chosen sampling interval does not have a pattern. Any pattern makes the sample not random. A random starting point must also be selected.



Systematic sampling

Systematic sampling is the selection of every k^{th} element from a sampling frame, where k, is the sampling interval.

$$k = \frac{number\ of\ the\ population}{number\ of\ the\ sample} = \frac{N}{n}$$

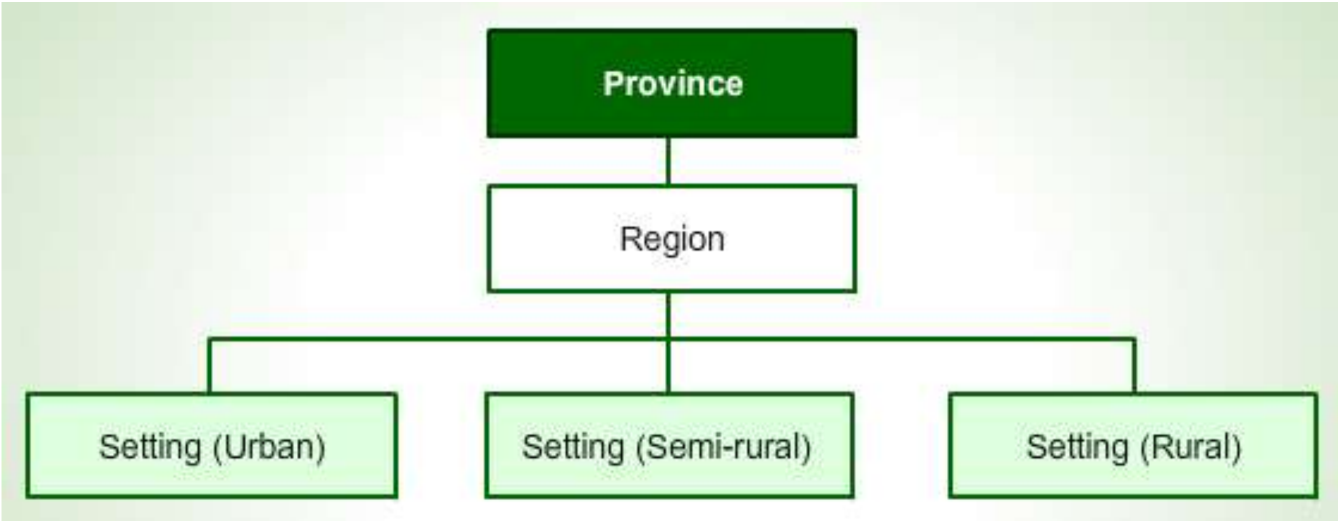
For instance, we have a population of N elements ordered and numbered from 1 to N, and we want to get a sample with n elements.

Here are steps to select a systematic random sample:

Step 1	Sample size is n.
Step 2	This population of N elements can be divided in n subsets, each of them with $k = \frac{N}{n}$ elements. In simple terms. Each subset has as many elements as the factor (k) indicates.
Step 3	We randomly choose a numbered element from the 1^{st} group, we call it x_0.
Step 4	We take the following elements: $x_0 + k$, $x_0 + 2k$, $x_0 + 3k$, $x_0 + 4k$ . . . In case k is not a natural number, we round to the closer one (lower), Some samples may have size $n - 1$.

Cluster Sampling

Cluster sampling is a sampling technique used when "natural" groupings are evident in the population. The total population is divided into these groups or clusters usually on a geographical basis, and a sample of the groups is selected. Then the required information is collected from the elements within each selected group. This may be done for every element in these groups, or a subsample of elements may be selected within each of these groups.



All items in the selected clusters can be used, or items can be chosen from a cluster using another probability sampling technique.

The main objective of cluster sampling is to reduce costs by increasing sampling efficiency.

Recap

In this lecture, you have learned that:

- There are two categorizes of sampling techniques:
    - Non statistical sampling techniques
    - Statistical sampling techniques
- A sample is called a simple random sample if each unit of the population has an equal chance of being selected for the sample
- Stratification is the process of grouping members of the population into relatively homogeneous subgroups before sampling
- In systematic sampling, each element in the population has a known and equal probability of selection
- Cluster sampling is a sampling technique used when "natural" groupings are evident in the population