

Data collection, Analysis and Inference

Subject Code: CPE-RPE,

May 2021,
SRM Univeristy-AP, Andhrapradesh

Lecture- 5: Sample size determination & sampling techniques:

Simple random sampling, Stratified sampling

- Aim: To implement the methods of sampling using Simple random sampling technique and Stratification.

Basics

- **Population**: It is the collection of a specified group of similar objects, individuals or of entities that have some common observation characteristics in them.
- **Sample**: A part of the population selected according to some rule or plan for drawing conclusion regarding population characteristics is called a sample.
- **Sampling**: Sampling means selecting a particular group or sample to represent the entire population.
- Two type of Sampling are there:
- 1. Probability Sampling, 2. Non-probability Sampling.

Probability Sampling:

It is a random sampling.

- Each individual of population has a fixed, known equal opportunity to be a part of the sample.
- This helps to reduce the possibility of bias.
- Different type of probability sampling methods:

1. Simple random sampling,
2. Systematic Sampling,
3. Stratified Random Sampling,
4. Cluster Random Sampling.

Non-probability Sampling:

- It is a non-random sampling.
- There is no specific chances of an individual to be a part of the sample.
- The possibility of bias are there.
- The selection of units depend on the subjective judgement of the researcher.

Simple random sampling

- **Definition:** Simple random sampling is a sampling technique where every item in the population has an even chance and likelihood of being selected in the sample.

Here the selection of items completely depends on chance or by probability and therefore this sampling technique is also sometimes known as **a method of chances**.

How it is done ?

- Step 1: Make a list.
- Step 2: Assign a sequential number.
- Step 3: Figure out what your sample size is going to be.
- Step 4: Use any of the following method to select the sample:
 1. Random number tables
 2. Random number generator software, or
 3. Method of lottery, etc.

Table A.1 Random Digits

94737	08225	35614	24826	88319	05595	58701	57365	74759
87259	85982	13296	89326	74863	99986	68558	06391	50248
63856	14016	18527	11634	96908	52146	53496	51730	03500
66612	54714	46783	61934	30258	61674	07471	67566	31635
30712	58582	05704	23172	86689	94834	99057	55832	21012
69607	24145	43886	86477	05317	30445	33456	34029	09603
37792	27282	94107	41967	21425	04743	42822	28111	09757
01488	56680	73847	64930	11108	44834	45390	86043	23973
66248	97697	38244	50918	55441	51217	54786	04940	50807
51453	03462	61157	65366	61130	26204	15016	85665	97714
92168	82530	19271	86999	96499	12765	20926	25282	39119
36463	07331	54590	00546	03337	41583	46439	40173	46455
47097	78780	04210	87084	44484	75377	57753	41415	09890
80400	45972	44111	99708	45935	03694	81421	60170	58457
94554	13863	88239	91624	00022	40471	78462	96265	55360
31567	53597	08490	73544	72573	30961	12282	97033	13676
07821	24759	47266	21747	72496	77755	50391	59554	31177
09056	10709	69314	11449	40531	02917	95878	74587	60906
19922	37025	80731	26179	16039	01518	82697	73227	13160
29923	02570	80164	36108	73689	26342	35712	49137	13482
29602	29464	99219	20308	82109	03898	82072	85199	13103
94135	94661	87724	88187	62191	70607	63099	40494	49069
87926	34092	34334	55064	43152	01610	03126	47312	59578
85039	19212	59160	83537	54414	19856	90527	21756	64783
66070	38480	74636	45095	86576	79337	39578	40851	53503
78166	82521	79261	12570	10930	47564	77869	16480	43972
94672	07912	26153	10531	12715	63142	88937	94466	31388
56406	70023	27734	22254	27685	67518	63966	33203	70803
67726	57805	94264	77009	08682	18784	47554	59869	66320
07516	45979	76735	46509	17696	67177	92600	55572	17245

- **Example 1:** Use a random digit table to select 5 numbers each of them are three digit number.

- **Example 2:** There are 41 students in a class, and 5 of them will be selected at random for estimating the average height.

Use a random digit table to select a simple random sample of 5 students.

Example 3: Willy runs a small company with 10 employees. He decides to pick a simple random sample of 3 employees to go on a business trip.

He numbers them 0-9 and uses the random digit table printed below to select the sample.

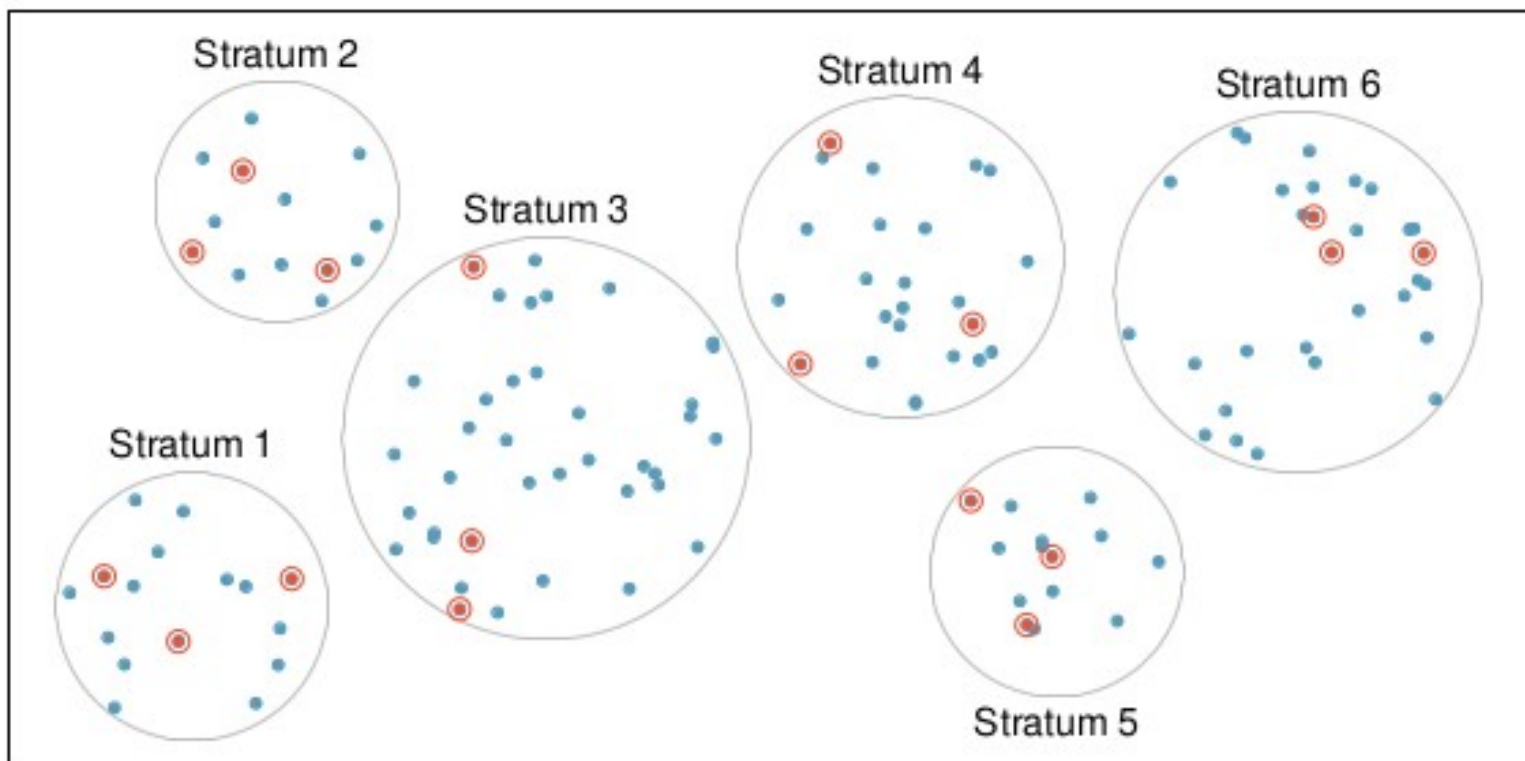
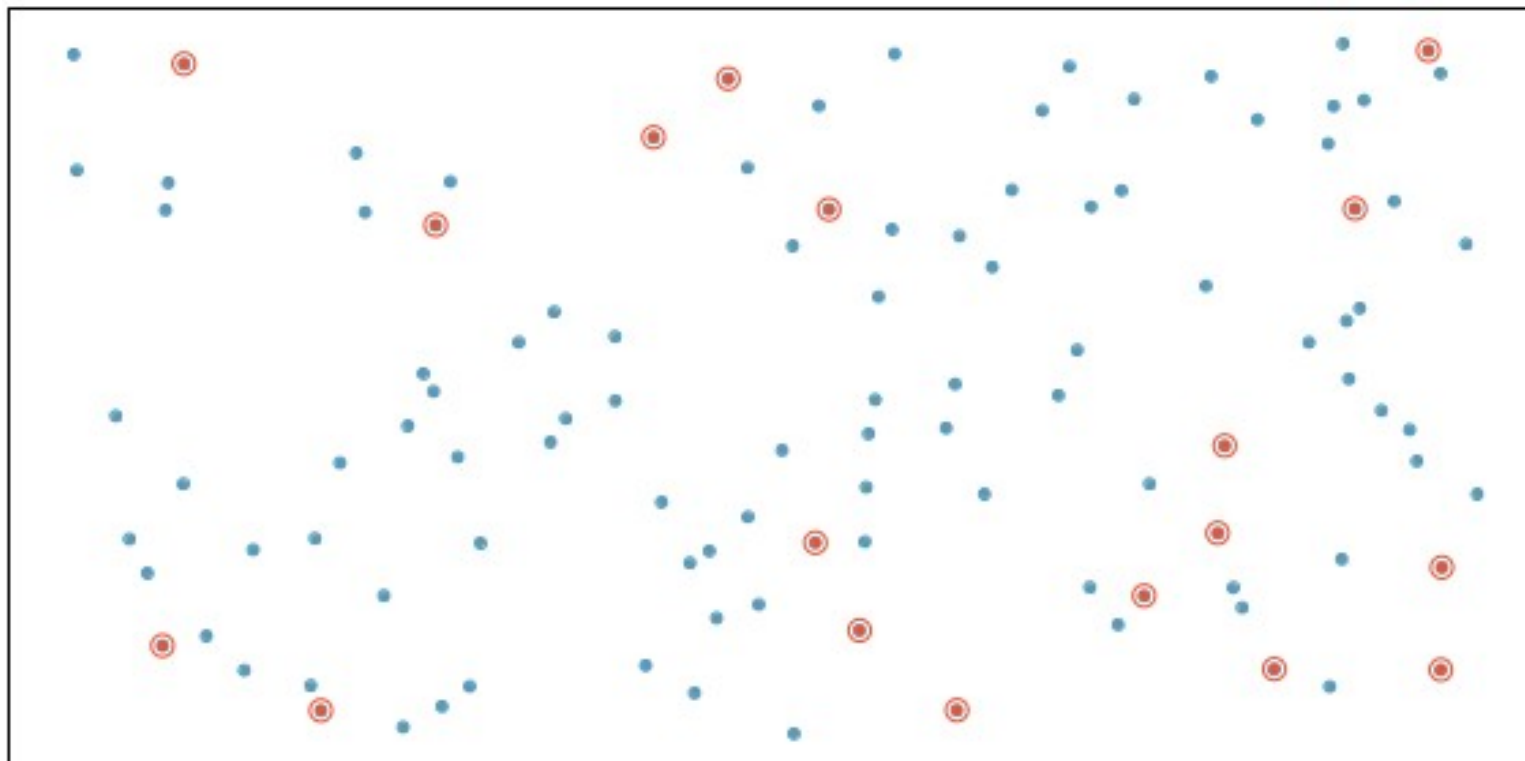
37374 75202 44753

- Which employees are in the sample?

A. 37,37,47 B. 37,47,52 C. 3,7,4 D. 3,7,3

Stratified sampling

- If the population is homogeneous with respect to the characteristic under study, then the method of SRS will yield a homogeneous sample, and in turn, the sample mean will serve as a good estimator of the population mean.
- The variance of the sample mean not only depends on the sample size but also on the population variance. To increase the precision one has to reduce the heterogeneity in the population.
- For this, **Stratified random sampling** is of great use.



The basic idea

....is to divide the whole heterogeneous population into sub-populations, such that the sampling units are homogeneous with respect to the characteristic under study **within that Sub-population,**

- And, heterogeneous with respect to the characteristic under study **among the sub-populations.**
- Such sub-populations are termed as **Strata.**
- Treat each Strata as a separate population and draw a sample by SRS from each stratum.

Note: 'Stratum' is singular and 'strata' is plural

Example: In order to find the average height of the students in a school of class 1 to class 12, the height varies a lot as the students in class 1 are of age around 6 years, and students in class 10 are of age around 16 years.

So one can divide all the students into different ***sub-populations or Strata*** such as

- **Students of class 1, 2 and 3: Stratum 1**
- **Students of class 4, 5 and 6: Stratum 2**
- **Students of class 7, 8 and 9: Stratum 3**
- **Students of class 10, 11 and 12: Stratum 4**

Now draw the samples by SRS from each of the Strata 1, 2, 3 and 4.

- **All the drawn samples combined together will constitute the final stratified sample for further analysis.**

Population= 1000

sample= 100

Urban

Rural

female

$p - 200$

20 %

$s - 20$

female

$p - 150$

15 %

$s - 15$

male

$p - 250$

25 %

$s - 25$

male

$p - 400$

40 %

$s = 40$

45 %

55 %

- END