

Get started

Open in app



Follow

616K Followers



You have **2** free member-only stories left this month. [Sign up for Medium and get an extra one](#)

8 Types of Sampling Techniques

Understanding Sampling Methods (Visuals and Code)



Prakhar Mishra Jun 10, 2021 · 5 min read ★

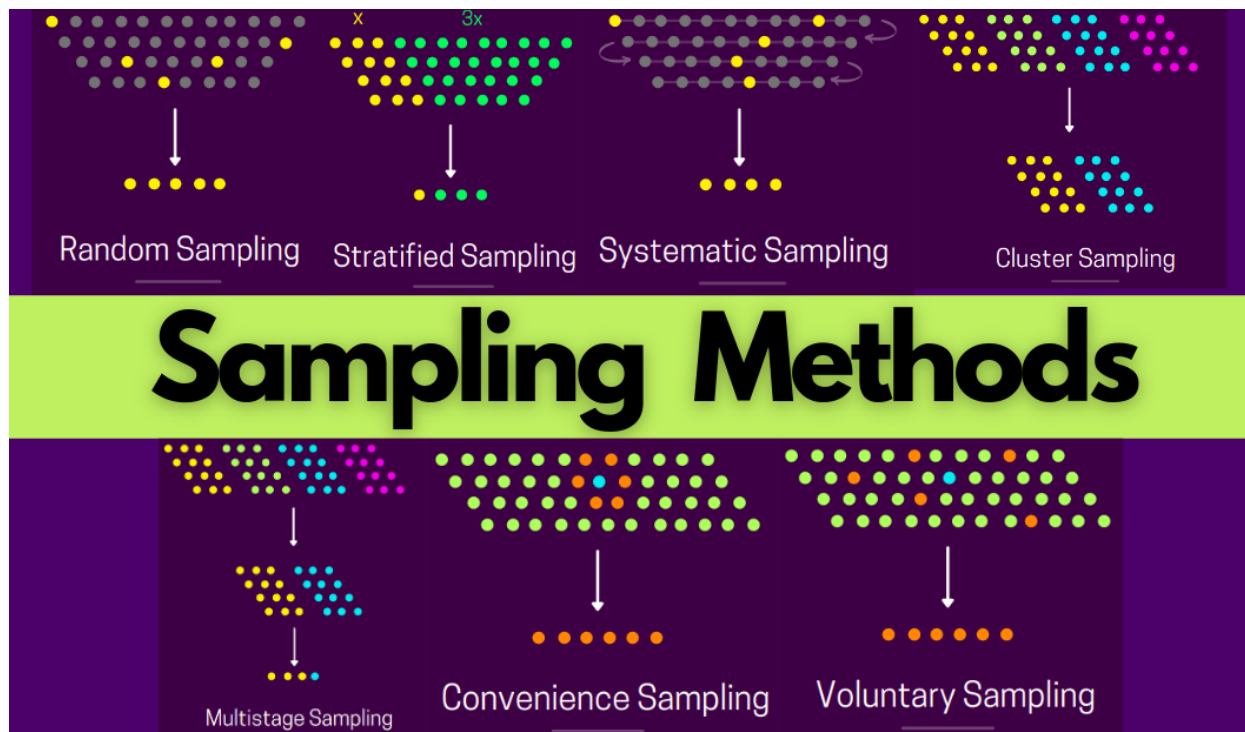


Image from Author

Sampling is the process of selecting a subset(*a predetermined number of observations*) from a larger population. It's a pretty common technique wherein, we run experiments and draw conclusions about the population, without the need of having to study the entire population. In this blog, we will go through two types of sampling methods:

1. **Probability Sampling** — Here we choose a sample based on the theory of probability.
2. **Non-Probability Sampling** — Here we choose a sample based on non-random criteria, and not every member of the population has a chance of being included.

Probability Sampling

Random Sampling

Under Random sampling, every element of the population has an equal probability of getting selected. Below fig. shows the pictorial view of the same — All the points collectively represent the entire population wherein every point has an equal chance of getting selected.



Random Sampling

You can implement it using python as shown below —

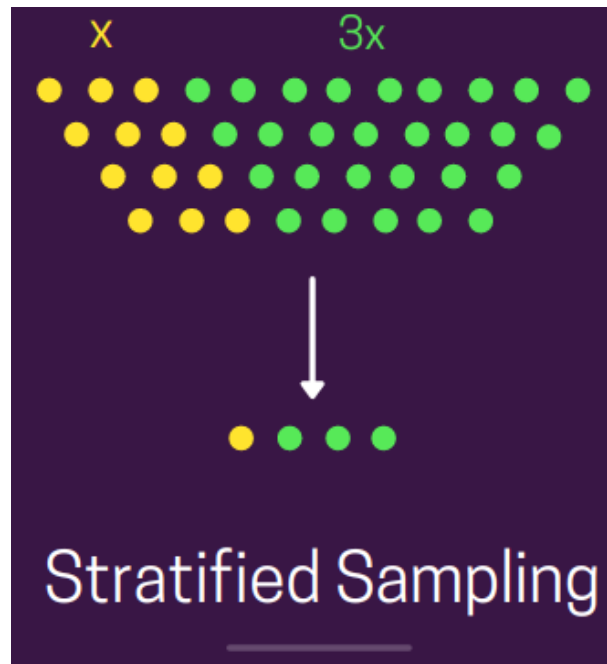
```
import random

population = 100
data = range(population)

print(random.sample(data,5))
> 4, 19, 82, 45, 41
```

Stratified Sampling

Under stratified sampling, we **group the entire population into subpopulations** by some common property. *For example — Class labels in a typical ML classification task.* We then randomly sample from those groups individually, such that the **groups are still maintained in the same ratio** as they were in the entire population. *Below fig. shows a pictorial view of the same — We have two groups with a count ratio of x and $4x$ based on the colour, we randomly sample from yellow and green sets separately and represent the final set in the same ratio of these groups.*



Stratified Sampling

You can implement it very easily using python sklearn lib. as shown below —

```
from sklearn.model_selection import train_test_split

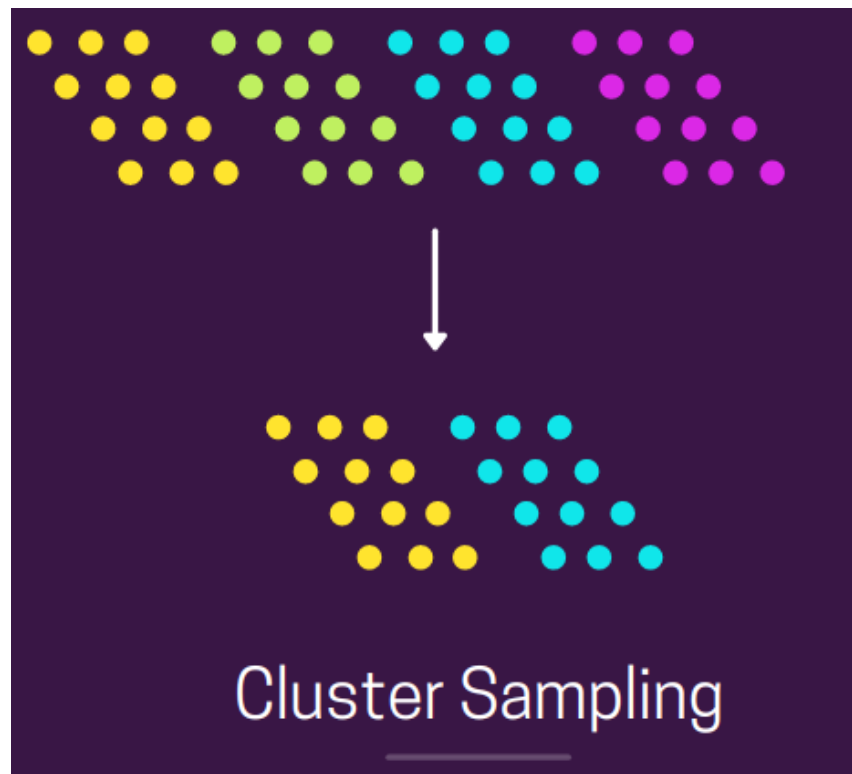
stratified_sample, _ = train_test_split(population, test_size=0.9,
stratify=population[['label']])
print (stratified_sample)
```

You can also implement it without the lib., [read this](#).

Cluster Sampling

In Cluster sampling, we **divide the entire population into subgroups**, wherein, each of those subgroups has similar characteristics to that of the population when considered in totality. Also, instead of sampling individuals, we **randomly select the entire**

subgroups. As can be seen in the below fig. that we had 4 clusters with similar properties (size and shape), we randomly select two clusters and treat them as samples.



Cluster Sampling

Real-Life example — Class of 120 students divided into groups of 12 for a common class project. Clustering parameters like (*Designation, Class, Topic*) are all similar over here as well.

You can implement it using python as shown below —

```
import numpy as np

clusters=5
pop_size = 100
sample_clusters=2

#assigning cluster ids sequentially from 1 to 5 on gap of 20
cluster_ids = np.repeat([range(1,clusters+1)], pop_size/clusters)

cluster_to_select = random.sample(set(cluster_ids), sample_clusters)

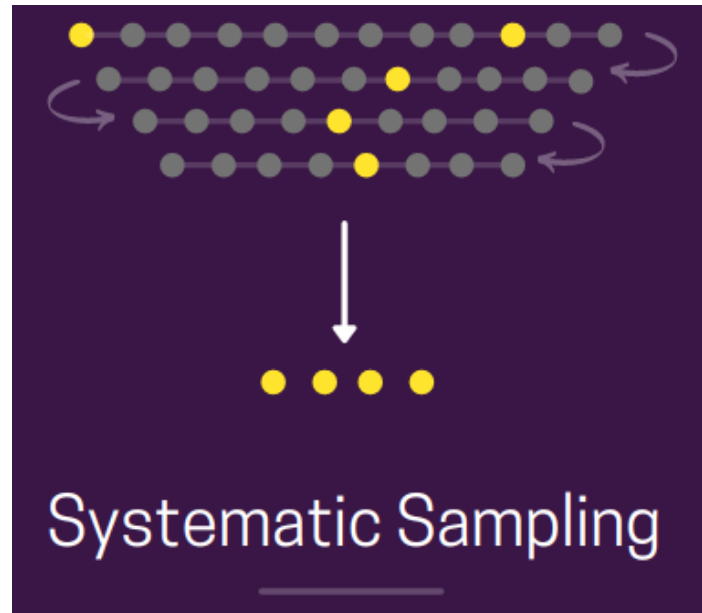
indexes = [i for i, x in enumerate(cluster_ids) if x in
cluster_to_select]

cluster_associated_elements = [el for idx, el in enumerate(range(1,
101)) if idx in indexes]
```

```
print (cluster_associated_elements)
```

Systematic Sampling

Systematic sampling is about sampling items from the population at **regular predefined intervals** (basically fixed and periodic intervals). For example — Every 5th element, 21st element and so on. This sampling method tends to be more effective than the vanilla random sampling method in general. Below fig. shows a pictorial view of the same — We sample every 9th and 7th element in order and then repeat this pattern.



Systematic Sampling

You can implement it using python as shown below —

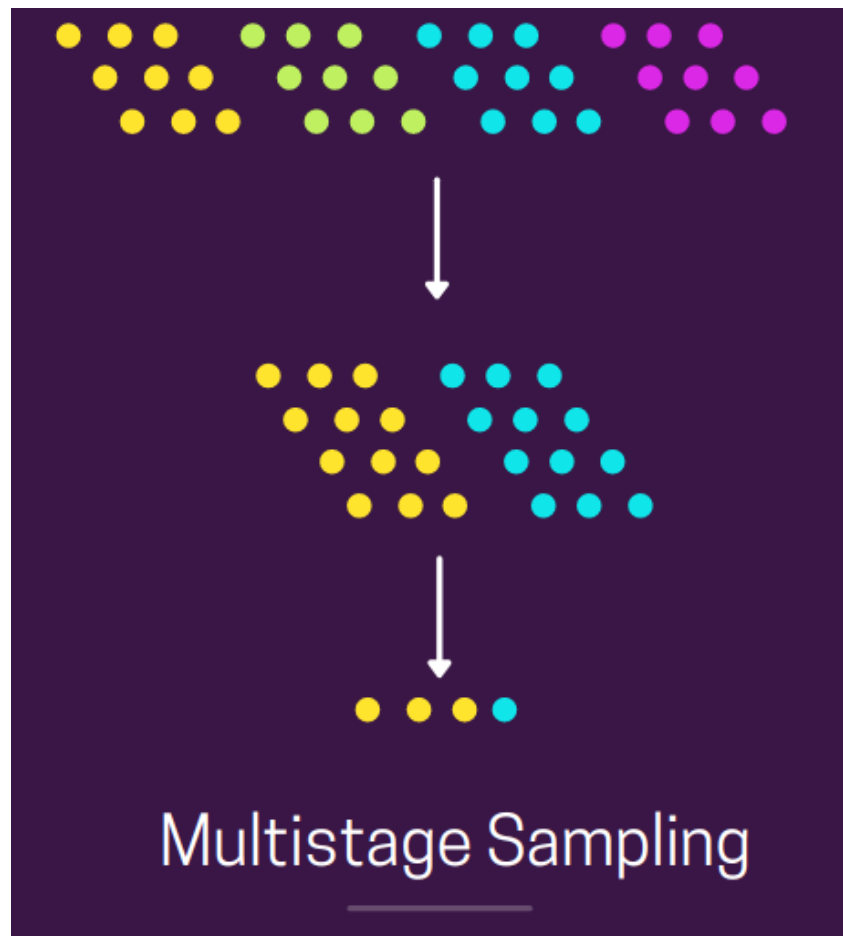
```
population = 100
step = 5

sample = [element for element in range(1, population, step)]
print (sample)
```

Multistage sampling

Under Multistage sampling, we **stack multiple sampling methods** one after the other. For example, at the first stage, cluster sampling can be used to choose clusters from the population and then we can perform random sampling to choose elements from each cluster to form the final set. Below fig. shows a pictorial view of the same —





Multi-stage Sampling

You can implement it using python as shown below —

```
import numpy as np

clusters=5
pop_size = 100
sample_clusters=2
sample_size=5

#assigning cluster ids sequentially from 1 to 5 on gap of 20
cluster_ids = np.repeat([range(1,clusters+1)], pop_size/clusters)

cluster_to_select = random.sample(set(cluster_ids), sample_clusters)

indexes = [i for i, x in enumerate(cluster_ids) if x in
cluster_to_select]

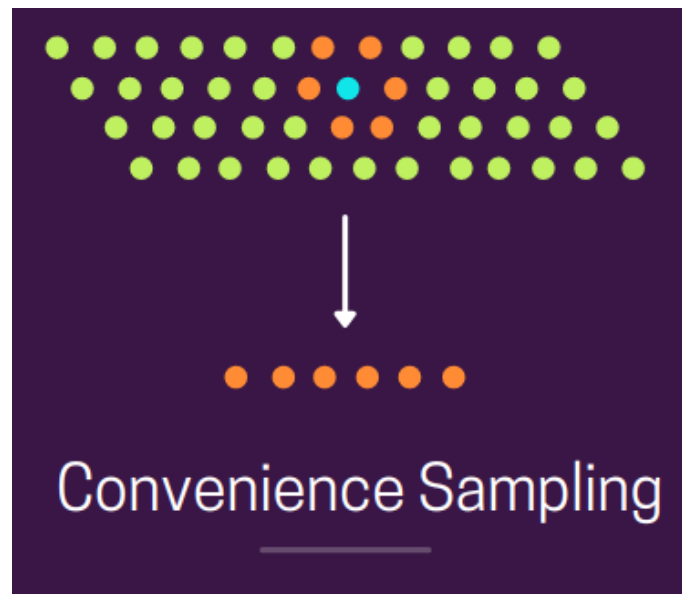
cluster_associated_elements = [el for idx, el in enumerate(range(1,
101)) if idx in indexes]

print (random.sample(cluster_associated_elements, sample_size))
```

Non-Probability Sampling

Convenience Sampling

Under convenience sampling, the researcher includes only those **individuals who are most accessible and available to participate in the study**. Below fig. shows the pictorial view of the same — Blue dot is the researcher and orange dots are the most accessible set of people in orange's vicinity.

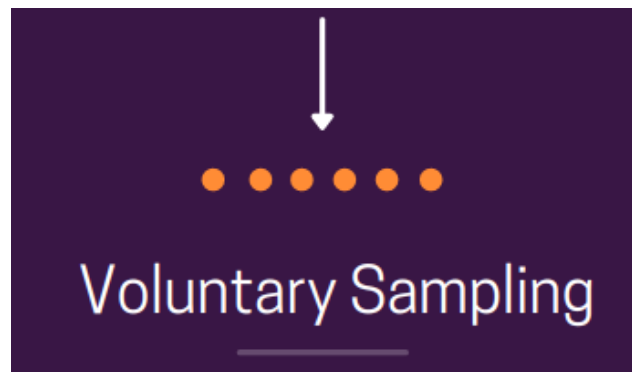


Convenience Sampling

Voluntary Sampling

Under Voluntary sampling, **interested people usually take part by themselves** by filling in some sort of survey forms. A good example of this is the youtube survey about “Have you seen any of these ads”, which has been recently shown a lot. Here, the **researcher who is conducting the survey has no right to choose anyone**. Below fig. shows the pictorial view of the same — Blue dot is the researcher, orange one's are those who voluntarily agreed to take part in the study.

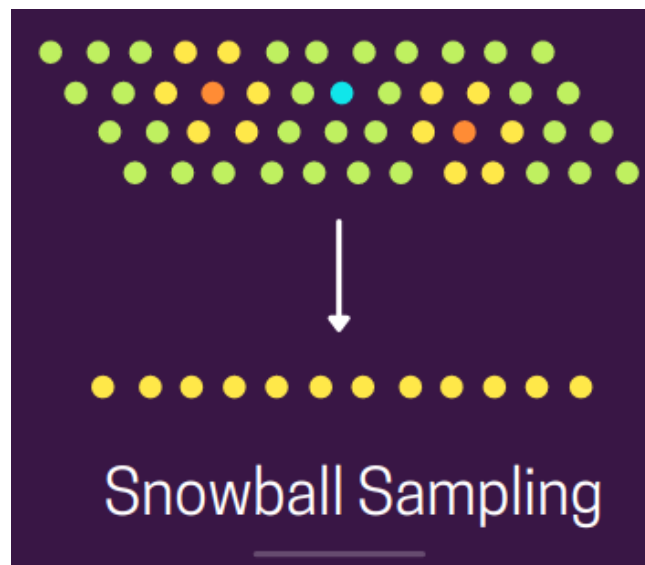




Voluntary Sampling

Snowball Sampling

Under Snowball sampling, the **final set is chosen via other participants**, i.e. The researcher asks other known contacts to find people who would like to participate in the study. *Below fig. shows the pictorial view of the same — Blue dot is the researcher, orange ones are known contacts(of the researcher), and yellow ones (orange's contacts) are other people that got ready to participate in the study.*



Snowball Sampling

Also if research papers interest you then you can [checkout some research paper summaries](https://towardsdatascience.com/8-types-of-sampling-techniques-b21adccd2124) that I have written.

I hope you enjoyed reading this. If you'd like to support me as a writer, consider signing up to [become a Medium member](#). It's just \$5 a month and you get unlimited access to Medium

So, that's it for this blog. Thank you for your time!

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

Get this newsletter

Statistics

Machine Learning

Probability

Data Science

Python

[About](#) [Write](#) [Help](#) [Legal](#)

Get the Medium app

