**Cluster Sampling(**[**https://www.scribbr.com/methodology/cluster-sampling/#:~:text=In%20cluster%20sampling%2C%20researchers%20divide,that%20are%20widely%20geographically%20dispersed**](https://www.scribbr.com/methodology/cluster-sampling/#:~:text=In%20cluster%20sampling%2C%20researchers%20divide,that%20are%20widely%20geographically%20dispersed)**. )**

* In cluster sampling, researchers divide a [population](https://www.scribbr.com/methodology/population-vs-sample/) into smaller groups known as **clusters**.  They then randomly select among these clusters to form a sample.
* Cluster sampling is a method of [probability sampling](https://www.scribbr.com/methodology/sampling-methods/#probability-sampling) that is often used to study large populations, particularly those that are widely geographically dispersed.

**Example:**

* You are interested in the average reading level of all the seventh-graders in your city.
* It would be very difficult to obtain a list of all seventh-graders and collect data from a random sample spread across the city. However, you can easily obtain a list of all schools and collect data from a subset of these. You thus decide to use the cluster sampling method.

Simplest form of cluster sampling is**single-stage cluster sampling**. It involves 4 key steps.

Step 1: Define your population :

Population In your reading program study, your population is all the seventh-graders in your city.

Step 2: Divide your sample into clusters

* The quality of your clusters and how well they **represent the larger population** determines the [validity](https://www.scribbr.com/methodology/internal-vs-external-validity/) of your results. Ideally, you would like for your clusters to meet the following criteria:
* Each cluster’s population should be as diverse as possible. You want every potential characteristic of the entire population to be represented in each cluster.
* Each cluster should have a similar distribution of characteristics as the distribution of the population as a whole.
* Taken together, the clusters should cover the entire population.
* There not be any overlap between clusters (i.e. the same people or units do not appear in more than one cluster).
* You cluster the seventh-graders by the school they attend. To cover the whole population, you need to include every school in the city. There is no overlap because each student attends only one school.

Problems :

Clusters often do not perfectly represent the population’s characteristics, which is why this method provides less statistical certainty than [simple random sampling](https://www.scribbr.com/methodology/simple-random-sampling/).

Because clusters are usually naturally occurring groups, such as schools, cities, or households, they are often more homogenous than the population as a whole. You should be aware of this when performing your study, as it might affect its validity.

Step 3: Randomly select clusters to use as your sample

* Cluster is itself a mini-representation of the larger population, randomly selecting and sampling from the clusters allows you to imitate simple random sampling, which in turn supports the validity of your results.
* Conversely, if the clusters are not representative, then random sampling will allow you to gather data on a diverse array of clusters, which should still provide you with an overview of the population as a whole
* You assign a number to each school and use a random number generator to select a random sample.
* You choose the number of clusters based on how large you want your sample size to be. This in turn is based on the estimated size of the entire seventh-grade population, your desired [confidence interval and confidence level](https://www.scribbr.com/statistics/confidence-interval/), and your best guess of the [standard deviation](https://www.scribbr.com/statistics/standard-deviation/) (a measure of how spread apart the values in a population are) of the reading levels of the seventh-graders.
* You then use a [sample size calculator](https://www.qualtrics.com/uk/experience-management/research/determine-sample-size/) to estimate the required sample size.

Step 4: Collect data from the sample

* Then conduct your study and collect data from every unit in the selected clusters.
* You test the reading levels of every seventh-grader in the schools that were randomly selected for your sample.