

# How can I perform cluster sampling in R, and what are some examples of its use?

Authored by  
**stats writer**

April 18, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I perform cluster sampling in R, and what are some examples of its use?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136759>

Cluster sampling is a method of selecting a sample from a population by dividing it into smaller groups or clusters and randomly selecting some clusters to be included in the sample. In R, cluster sampling can be performed using the "survey" package. This package provides functions for creating a sampling frame, selecting clusters, and estimating population parameters using the selected clusters. Some examples of cluster sampling in R include selecting households from different neighborhoods to survey their opinions on a certain topic, selecting schools from different districts to study their academic performance, and selecting hospitals from different regions to analyze their patient satisfaction rates. Cluster sampling is useful in situations where the population is geographically dispersed or when it is not feasible to obtain a complete list of all individuals in the population. It also allows for efficient and cost-effective sampling, making it a popular method in various research studies and surveys.

## Cluster Sampling in R (With Examples)

**Researchers often take samples from a population and use the data from the sample to draw conclusions about the population as a whole.**

**One commonly used sampling method is cluster sampling, in which a population is split into clusters and all members of *some* clusters are chosen to be included in the sample.**

**This tutorial explains how to perform cluster sampling in R.**

### **Example: Cluster Sampling in R**

**Suppose a company that gives city tours wants to survey its customers. Out of ten tours they give one**

day, they randomly select four tours and ask every customer to rate their experience on a scale of 1 to 10.

The following code shows how to create a fake data frame in R to work with:

```
#make this example reproducible  
set.seed(1)#create data frame  
df <- data.frame(tour = rep(1:10, each=20),  
experience = rnorm(200, mean=7, sd=1))
```

```
#view first six rows of data frame  
head(df)
```

```
tour experience
```

```
1 1 6.373546
```

```
2 1 7.183643
```

```
3 1 6.164371
```

```
4 1 8.595281
```

```
5 1 7.329508
```

```
6 1 6.179532
```

And the following code shows how obtain a sample of customers by randomly selecting four tours and including every member in those tours in the sample:

```
#randomly choose 4 tour groups out of the 10  
clusters <- sample(unique(df$tour), size=4, replace=F)  
  
#define sample as all members who belong to one of  
the 4 tour groups  
cluster_sample <- df  
  
#view how many customers came from each tour  
table(cluster_sample$tour)  
  
2 7 8 10  
20 20 20 20
```

From the output we can see that:

20 customers from tour group #2 were included in the sample. 20 customers from tour group #7 were included in the sample. 20 customers from tour group #8 were included in the sample. 20 customers from tour group #10 were included in the sample.

Thus, this sample is composed of 80 total customers that came from 4 different tour groups.

## How to Use %in% Operator in R

## **Understanding Different Types of Sampling Methods**

### **Stratified Sampling in R**

### **Systematic Sampling in R**

ARABPSYCHOLOGY.COM