

How can I select random samples in R, and can you provide some examples?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I select random samples in R, and can you provide some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137654>

Random sampling is a statistical method used to select a subset of data from a larger population, in order to make generalizations about the population as a whole. In R, there are various functions and packages available for performing random sampling. These include the base function "sample()" and the package "randomizr". To select random samples in R, you can specify the size of the sample and the population from which it should be drawn. Additionally, you can set parameters such as with or without replacement, and the probability of each element being selected. Here are some examples of random sampling in R:

1. Using the "sample()" function to select a random sample of 50 observations from a vector of numbers:

```
sample(1:100, size = 50)
```

2. Using the "sample()" function to select a random sample of 100 observations from a data frame:

```
sample(data_frame, size = 100, replace = FALSE)
```

3. Using the "sample()" function to select a random sample of 20 observations with a probability of 0.5 for each element being selected from a vector of letters:

```
sample(letters, size = 20, prob = c(0.5, 0.5), replace = TRUE)
```

4. Using the "randomizr" package to perform stratified random sampling, where the sample is drawn from different levels of a categorical variable:

```
strata_sample(data_frame, strata = categorical_variable, size = 50)
```

Overall, selecting random samples in R allows for unbiased and representative samples to be obtained, providing a solid foundation for statistical analysis and inference.

Select Random Samples in R (With Examples)

To select a random sample in R we can use the **sample()** function, which uses the following syntax:

sample(x, size, replace = FALSE, prob = NULL)

where:

x: A vector of elements from which to choose.size:

Sample size.replace: Whether to sample with replacement or not. Default is FALSE.**prob:** Vector of probability weights for obtaining elements from vector. Default is NULL.

This tutorial explains how to use this function to select a random sample in R from both a vector and a data frame.

Example 1: Random Sample from a Vector

The following code shows how to select a random sample from a vector without replacement:

```
#create vector of data
data <- c(1, 3, 5, 6, 7, 8, 10, 11, 12, 14)

#select random sample of 5 elements without
replacement
sample(x=data, size=5)

10 12 5 14 7
```

The following code shows how to select a random sample from a vector with replacement:

#create vector of data

```
data <- c(1, 3, 5, 6, 7, 8, 10, 11, 12, 14)
```

#select random sample of 5 elements with replacement

```
sample(x=data, size=5, replace=TRUE)
```

```
12 1 1 6 14
```

Example 2: Random Sample from a Data Frame

The following code shows how to select a random sample from a data frame:

#create data frame

```
df <- data.frame(x=c(3, 5, 6, 6, 8, 12, 14),  
y=c(12, 6, 4, 23, 25, 8, 9),  
z=c(2, 7, 8, 8, 15, 17, 29))
```

#view data frame

```
df
```

```
x y z
```

```
1 3 12 2
```

```
2 5 6 7
```

```
3 6 4 8
```

```
4 6 23 8
```

5 8 25 15

6 12 8 17

7 14 9 29

```
#select random sample of three rows from data frame  
rand_df <- df
```

```
#display randomly selected rows  
rand_df
```

x y z

4 6 23 8

7 14 9 29

1 3 12 2

Here's what's happening in this bit of code:

1. To select a subset of a data frame in R, we use the following syntax: `df`
2. In the code above, we randomly select a sample of 3 rows from the data frame and *all* columns.
3. The end result is a subset of the data frame with 3 randomly selected rows.

In order to replicate the results of some analysis, be sure to use `set.seed(some number)` so that the `sample()` function chooses the same random sample each time. For example:

```
#make this example reproducible  
set.seed(23)
```

```
#create data frame
```

```
df <- data.frame(x=c(3, 5, 6, 6, 8, 12, 14),  
y=c(12, 6, 4, 23, 25, 8, 9),  
z=c(2, 7, 8, 8, 15, 17, 29))
```

```
#select random sample of three rows from data frame  
rand_df <- df
```

```
#display randomly selected rows  
rand_df
```

```
x y z  
5 8 25 15  
2 5 6 7  
6 12 8 17
```

Each time you run the above code, the same 3 rows of

the data frame will be selected each time.

Stratified Sampling in R (With Examples)

Systematic Sampling in R (With Examples)

Cluster Sampling in R (With Examples)

ARABPSYCHOLOGY.COM