

Simple Random Sampling: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2020

Syntax

SIMPLE RANDOM SAMPLING

- Simple random sampling of a vector, sample size 10:

```
sample_10 <- sample(x, size = 10)
```

- Simple random sampling of a vector, sample size 500:

```
sample_500 <- sample(x, size = 500)
```

REPRODUCIBLE RANDOM SAMPLES

- Make the generation of random numbers reproducible:

```
set.seed(1)

sample_10 <- sample(x, size = 10)

set.seed(n)

sample_10 <- sample(x, size = 10)
```

USING FUNCTIONALS TO PERFORM SIMPLE RANDOM SAMPLING

- Replicate numerous random samples in a reproducible way:

```
set.seed(1)

replicate(n = 100,
          expr = sample(df$col, size = 10))
```

Concepts

- The set of *all* individuals relevant to a particular statistical question is called a **population**. A smaller group selected from a population is called a **sample**. When we select a smaller group from a population, we do **sampling**.

- A **parameter** is a metric specific to a population, and a **statistic** is a metric specific to a sample. The difference between a statistic and its corresponding parameter is called **sampling error**. If the sampling error is low, then the sample is **representative**.
- **Simple random sampling** is the process of generating random numbers and using them to select a few sample units from the population.

Resources

- [The Wikipedia entry](#) on sampling.
- [The Wikipedia entry](#) on samples.
- [The Wikipedia entry](#) on populations.



Takeaways by Dataquest Labs, Inc. - All rights reserved © 2020