

What is stratified sampling and how is it used in R? Can you provide examples of stratified sampling in R?

Authored by
stats writer

April 18, 2024

RECOMMENDED CITATION

stats writer (2024). *What is stratified sampling and how is it used in R? Can you provide examples of stratified sampling in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136750>

Stratified sampling is a statistical method used to select a sample from a population in a way that ensures representation of different subgroups or strata within the population. It involves dividing the population into smaller, homogeneous groups and then selecting a random sample from each group.

In R, stratified sampling can be performed using the "strata" function from the "survey" package. This function allows users to specify the variables used for stratification and the desired sample size from each stratum. The resulting sample will be representative of the entire population, as well as each individual stratum.

For example, if we want to conduct a survey on the preferences of students in a university, we can use stratified sampling by dividing the population into different majors (strata) and selecting a random sample of students from each major. This ensures that the sample reflects the diversity of the student population and provides more accurate results.

Another example could be conducting a market research study on a specific product in a city. The city can be divided into different neighborhoods (strata) and a random sample of households can be selected from each neighborhood to ensure representation of all areas within the city. This allows for a more accurate understanding of the market and consumer behavior.

Stratified 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 stratified random sampling, in which a population is split into groups and a certain number of members from each group are randomly selected to be included in the sample.

This tutorial explains how to perform stratified random

sampling in R.

Example: Stratified Sampling in R

A high school is composed of 400 students who are either Freshman, Sophomores, Juniors, or Seniors. Suppose we'd like to take a stratified sample of 40 students such that 10 students from each grade are included in the sample.

The following code shows how to generate a sample data frame of 400 students:

```
#make this example reproducible  
set.seed(1)
```

```
#create data frame
```

```
df <- data.frame(grade = rep(c('Freshman', 'Sophomore',  
'Junior', 'Senior'), each=100),
```

```
gpa = rnorm(400, mean=85, sd=3))
```

```
#view first six rows of data frame
```

```
head(df)
```

```
grade gpa
```

```
1 Freshman 83.12064
```

2 Freshman 85.55093
3 Freshman 82.49311
4 Freshman 89.78584
5 Freshman 85.98852
6 Freshman 82.53859

Stratified Sampling Using Number of Rows

The following code shows how to use the `group_by()` and `sample_n()` functions from the dplyr package to obtain a stratified random sample of 40 total students with 10 students from each grade:

```
library(dplyr)
```

```
#obtain stratified sample
```

```
strat_sample <- df %>%
```

```
group_by(grade) %>%
```

```
sample_n(size=10)
```

```
#find frequency of students from each grade
```

```
table(strat_sample$grade)
```

Freshman Junior Senior Sophomore

10 10 10 10

Stratified Sampling Using Fraction of Rows

The following code shows how to use the `group_by()` and `sample_frac()` functions from the **dplyr** package to obtain a stratified random sample in which we randomly select 15% of students from each grade:

```
library(dplyr)
```

```
#obtain stratified sample
```

```
strat_sample <- df %>%
```

```
group_by(grade) %>%
```

```
sample_frac(size=.15)
```

```
#find frequency of students from each grade
```

```
table(strat_sample$grade)
```

```
Freshman Junior Senior Sophomore
```

```
15 15 15 15
```

Types of Sampling Methods

Cluster Sampling in R

Systematic Sampling in R