

# How do you calculate sampling distributions in R?

Authored by  
**stats writer**

April 22, 2024

## RECOMMENDED CITATION

stats writer (2024). *How do you calculate sampling distributions in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138133>

Sampling distribution refers to the probability distribution of a statistic, such as mean or standard deviation, calculated from a random sample of a population. In R, the process of calculating sampling distributions involves generating random samples from a population, calculating the desired statistic for each sample, and then analyzing the distribution of these statistics. This can be done using built-in functions and packages in R, such as the "sample" function and the "boot" package. By simulating multiple random samples and analyzing their statistics, one can estimate the true population parameters and understand the variability of the statistic. This process is commonly used in statistical inference and hypothesis testing.

## Calculate Sampling Distributions in R

**A sampling distribution is a probability distribution of a certain statistic based on many random samples from a single population.**

**This tutorial explains how to do the following with sampling distributions in R:**

**Generate a sampling distribution. Visualize the sampling distribution. Calculate the mean and standard deviation of the sampling distribution. Calculate probabilities regarding the sampling distribution.**

**Generate a Sampling Distribution in R**

**The following code shows how to generate a sampling distribution in R:**

**#make this example reproducible**

```
set.seed(0)
```

```
#define number of samples
```

```
n = 10000
```

```
#create empty vector of length n
```

```
sample_means = rep(NA, n)
```

```
#fill empty vector with means
```

```
for(i in 1:n){
```

```
sample_means = mean(rnorm(20, mean=5.3, sd=9))
```

```
}
```

```
#view first six sample means
```

```
head(sample_means)
```

```
5.283992 6.304845 4.259583 3.915274 7.756386 4.532656
```

In this example we used the **rnorm()** function to calculate the mean of 10,000 samples in which each sample size was 20 and was generated from a normal distribution with a mean of 5.3 and standard deviation of 9.

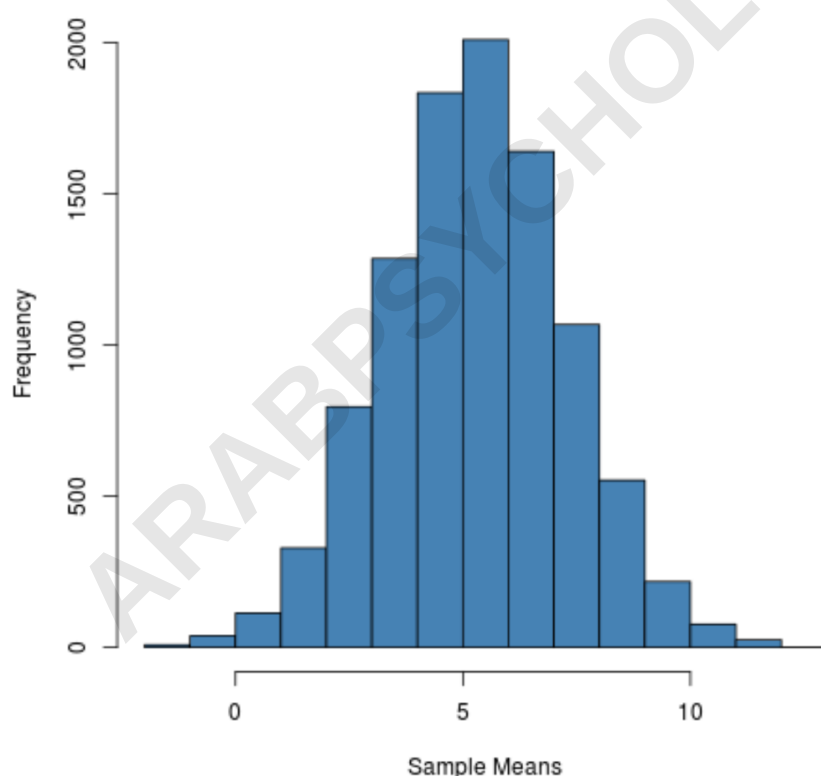
We can see that the first sample had a mean of 5.283992, the second sample had a mean of 6.304845,

and so on.

### Visualize the Sampling Distribution

The following code shows how to create a simple histogram to visualize the sampling distribution:

```
#create histogram to visualize the sampling distribution  
hist(sample_means, main = "", xlab = "Sample Means",  
col = "steelblue")
```



We can see that the sampling distribution is bell-shaped with a peak near the value 5.

From the tails of the distribution, however, we can see that some samples had means greater than 10 and some had means less than 0.

Find the Mean & Standard Deviation

The following code shows how to calculate the mean and standard deviation of the sampling distribution:

```
#mean of sampling distribution
```

```
mean(sample_means)
```

```
5.287195
```

```
#standard deviation of sampling distribution
```

```
sd(sample_means)
```

```
2.00224
```

T

And theoretically the standard deviation of the sampling distribution should be equal to  $s/\sqrt{n}$ , which would be  $9 / \sqrt{20} = 2.012$ . We can see that the actual standard deviation of the sampling distribution is 2.00224, which is close to 2.012.

## Calculate Probabilities

The following code shows how to calculate the probability of obtaining a certain value for a sample mean, based on a population mean, population standard deviation, and sample size.

```
#calculate probability that sample mean is less than or  
equal to 6
```

```
sum(sample_means <= 6) / length(sample_means)
```

In this particular example, we find the probability that the sample mean is less than or equal to 6, given that the population mean is 5.3, the population standard deviation is 9, and the sample size is 20 is 0.6417.

This is very close to the probability calculated by the **Sampling Distribution Calculator**:

$\mu$  (population mean)

$\sigma$  (population standard deviation)

$n$  (sample size)

$X$  (random variable)

CALCULATE

$P(\bar{X} \leq 6): 0.63602$

$P(\bar{X} \geq 6): 0.36398$

### The Complete Code

The complete R code used in this example is shown below:

```
#make this example reproducible  
set.seed(0)
```

```
#define number of samples
```

**n = 10000**

**#create empty vector of length n**

**sample\_means = rep(NA, n)**

**#fill empty vector with means**

**for(i in 1:n){**

**sample\_means = mean(rnorm(20, mean=5.3, sd=9))**

**}**

**#view first six sample means**

**head(sample\_means)**

**#create histogram to visualize the sampling distribution**

**hist(sample\_means, main = "", xlab = "Sample Means",  
col = "steelblue")**

**#mean of sampling distribution**

**mean(sample\_means)**

**#standard deviation of sampling distribution**

**sd(sample\_means)**

**#calculate probability that sample mean is less than or  
equal to 6**

**sum(sample\_means <= 6) / length(sample\_means)**

## **An Introduction to Sampling Distributions**

## **Sampling Distribution Calculator**

## **An Introduction to the Central Limit Theorem**

ARABPSYCHOLOGY.COM