

How can we use statistics to estimate population means?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How can we use statistics to estimate population means?*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159180>

Statistics is the study of collecting, analyzing, interpreting, and presenting data in order to make informed decisions and draw conclusions. One of the key uses of statistics is to estimate population means. This involves using sample data to make predictions and draw inferences about the characteristics of a larger population. By collecting a representative sample from the population and using statistical techniques, such as hypothesis testing and confidence intervals, we can estimate the true mean of the population with a certain level of confidence. This provides valuable information for decision making and understanding the characteristics of a population without having to survey every individual. Overall, statistics allows us to make reliable estimates of population means and gain a better understanding of the world around us.

Statistics - Estimating Population Means

A population mean is an average of a numerical population variable.

Confidence intervals are used to estimate population means.

Estimating Population Mean

A statistic from a sample is used to estimate a parameter of the population.

The most likely value for a parameter is the **point estimate**.

Additionally, we can calculate a **lower bound** and an **upper bound** for the estimated parameter.

The **margin of error** is the difference between the lower and upper bounds from the point estimate.

Together, the lower and upper bounds define a **confidence interval**.

Calculating a Confidence Interval

The following steps are used to calculate a confidence interval:

Check the conditions
Find the point estimate
Decide the confidence level
Calculate the margin of error
Calculate the confidence interval

For example:

Population: Nobel Prize winners
Variable: Age when they received the Nobel Prize

We can take a sample and calculate the mean and the standard deviation of that sample.

The sample data is used to make an estimation of the average age of **all** the Nobel Prize winners.

By randomly selecting 30 Nobel Prize winners we could find that:

The mean age in the sample is 62.1

The standard deviation of age in the sample is 13.46

From this data we can calculate a confidence interval with the steps below.

1. Checking the Conditions

The conditions for calculating a confidence interval for a mean are:

The sample is randomly selected And either:

The population data is normally distributed Sample size is large enough

A moderately large sample size, like 30, is typically large enough.

In the example, the sample size was 30 and it was randomly selected, so the conditions are fulfilled.

Note: Checking if the data is normally distributed can be done with specialized statistical tests.

2. Finding the Point Estimate

The point estimate is the sample mean (\bar{x}).

The formula for calculating the sample mean is the sum of all the values ($\sum x_{\{i\}}$) divided by the sample size (n):

$$\bar{x} = \frac{\sum x_{\{i\}}}{n}$$

In our example, the mean age was 62.1 in the sample.

3. Deciding the Confidence Level

The confidence level is expressed with a percentage or a decimal number.

For example, if the confidence level is 95% or 0.95:

The remaining probability (α) is then: 5%, or $1 - 0.95 = 0.05$.

Commonly used confidence levels are:

90% with $(\alpha) = 0.1$ 95% with $(\alpha) = 0.05$ 99% with $(\alpha) = 0.01$

Note: A 95% confidence level means that if we take 100 different samples and make confidence intervals for each:

The true parameter will be inside the confidence interval 95 out of those 100 times.

We use the student's t-distribution to find the **margin of error** for the confidence interval.

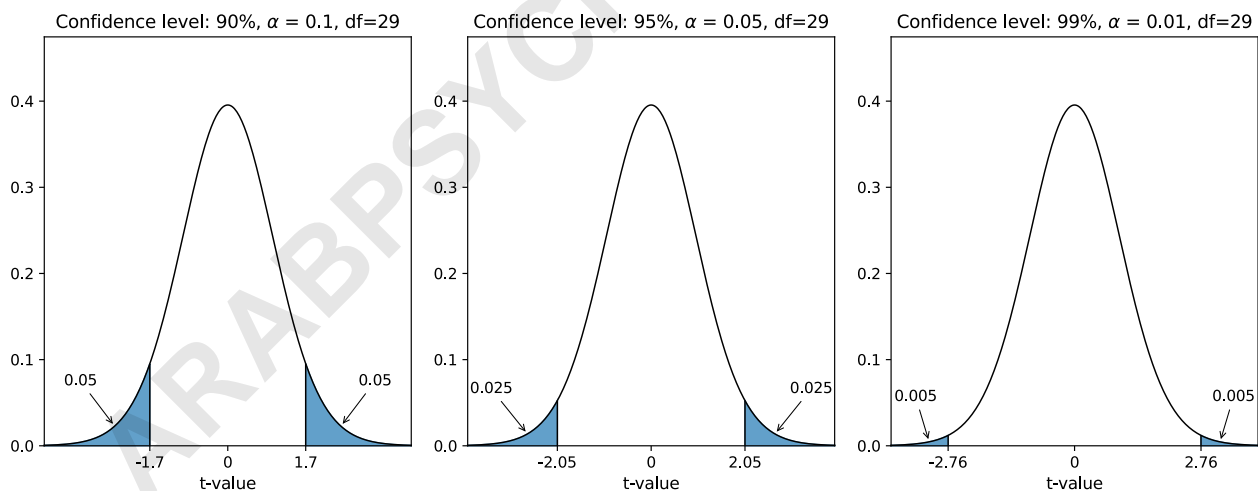
The t-distribution is adjusted for the sample size with 'degrees of freedom' (df).

The degrees of freedom is the sample size (n) - 1, so in this example it is $30 - 1 = 29$

The remaining probabilities ((alpha)) are divided in two so that half is in each tail area of the distribution.

The values on the t-value axis that separate the tails area from the middle are called **critical t-values**.

Below are graphs of the standard normal distribution showing the tail areas ((alpha)) for different confidence levels at 29 degrees of freedom (df).



4. Calculating the Margin of Error

The margin of error is the difference between the point estimate and the lower and upper bounds.

The margin of error ((E)) for a proportion is calculated with a critical t-value and the **standard error**:

$$E = t_{\alpha/2}(df) \cdot \frac{s}{\sqrt{n}}$$

The critical t-value ($t_{\alpha/2}(df)$) is calculated from the standard normal distribution and the confidence level.

The standard error ($\frac{s}{\sqrt{n}}$) is calculated from the sample standard deviation ((s)) and the sample size ((n)).

In our example with a sample standard deviation ((s)) of 13.46 and sample size of 30 the standard error is:

$$\frac{s}{\sqrt{n}} = \frac{13.46}{\sqrt{30}} \approx \frac{13.46}{5.477} = \underline{2.458}$$

If we choose 95% as the confidence level, the (α) is 0.05.

So we need to find the critical t-value ($t_{\{0.05/2\}(29)} = t_{\{0.025\}(29)}$)

The critical t-value can be found using a t-table or with a programming language function:

Example

With Python use the Scipy Stats library `t.ppf()` function find the t-value for an $(\alpha)/2 = 0.025$ and 29 degrees of freedom.

```
import scipy.stats as stats
```

```
print(stats.t.ppf(1-0.025, 29))
```

Example

With R use the built-in `qt()` function to find the t-value for an $(\alpha)/2 = 0.025$ and 29 degrees of freedom.

```
qt(1-0.025, 29)
```

Using either method we can find that the critical t-value ($t_{\alpha/2}(df)$) is (approx 2.05)

The standard error ($\frac{s}{\sqrt{n}}$) was (approx 2.458)

So the margin of error ((E)) is:

$$E = t_{\alpha/2}(df) \cdot \frac{s}{\sqrt{n}} \approx 2.05 \cdot 2.458 = \underline{5.0389}$$

5. Calculate the Confidence Interval

The lower and upper bounds of the confidence interval are found by subtracting and adding the margin of error (E) from the point estimate (\bar{x}).

In our example the point estimate was 0.2 and the margin of error was 0.143, then:

The lower bound is:

$$(\bar{x} - E = 62.1 - 5.0389 \approx \underline{57.06})$$

The upper bound is:

$$(\bar{x} + E = 62.1 + 5.0389 \approx \underline{67.14})$$

The confidence interval is:

()

And we can summarize the confidence interval by stating:

The **95%** confidence interval for the mean age of Nobel Prize winners is between **57.06 and 67.14 years**

Calculating a Confidence Interval with Programming

A confidence interval can be calculated with many programming languages.

Using software and programming to calculate statistics is more common for bigger sets of data, as calculating manually becomes difficult.

Note: The results from using the programming code will be more accurate because of rounding of values when calculating by hand.

Example

With Python use the `scipy` and `math` libraries to calculate the confidence interval for an estimated proportion.

Here, the sample size is 30, sample mean is 62.1 and sample standard deviation is 13.46.

```
import scipy.stats as stats
```

```
import math

# Specify sample mean (x_bar), sample standard deviation (s), sample size (n) and confidence
level

x_bar = 62.1

s = 13.46

n = 30

confidence_level = 0.95

# Calculate alpha, degrees of freedom (df), the critical t-value, and the margin of error

alpha = (1-confidence_level)

df = n - 1

standard_error = s/math.sqrt(n)

critical_t = stats.t.ppf(1-alpha/2, df)

margin_of_error = critical_t * standard_error

# Calculate the lower and upper bound of the confidence interval

lower_bound = x_bar - margin_of_error

upper_bound = x_bar + margin_of_error

# Print the results

print("Critical t-value: {:.3f}".format(critical_t))

print("Margin of Error: {:.3f}".format(margin_of_error))

print("Confidence Interval: {}".format(lower_bound,upper_bound))

print("The {:.1%} confidence interval for the population mean is:".format(confidence_level))
```

```
print("between {:.3f} and {:.3f}".format(lower_bound,upper_bound))
```

Example

R can use built-in math and statistics functions to calculate the confidence interval for an estimated proportion.

Here, the sample size is 30, sample mean is 62.1 and sample standard deviation is 13.46.

```
# Specify sample mean (x_bar), sample standard deviation (s), sample size (n) and confidence level
```

```
x_bar = 62.1
```

```
s = 13.46
```

```
n = 30
```

```
confidence_level = 0.95
```

```
# Calculate alpha, degrees of freedom (df), the critical t-value, and the margin of error
```

```
alpha = (1-confidence_level)
```

```
df = n - 1
```

```
standard_error = s/sqrt(n)
```

```
critical_t = qt(1-alpha/2, 29)
```

```
margin_of_error = critical_t * standard_error
```

```
# Calculate the lower and upper bound of the confidence interval
```

```
lower_bound = x_bar - margin_of_error
```

```
upper_bound = x_bar + margin_of_error
```

```
# Print the results
```

```
sprintf("Critical t-value: %0.3f", critical_t)
```

```
sprintf("Margin of Error: %0.3f", margin_of_error)
```

```
sprintf("Confidence Interval: ", lower_bound, upper_bound)
```

```
sprintf("The %0.1f%% confidence interval for the population mean is:", confidence_level*100)
```

```
sprintf("between %0.4f and %0.4f", lower_bound, upper_bound)
```

Note: R also has a built in function for calculating a confidence interval for a population mean.

Example

R can use the built-in `t.test()` function to calculate the confidence interval for an estimated mean.

Here, the sample is 30 randomly generated values with a mean of 60 and standard deviation is 12.5 using the `rnorm()` function to generate the sample.

```
# Specify sample size (n) and confidence level
```

```
n = 30
```

```
confidence_level = 0.95
```

```
# Set random seed and generate sample data with mean of 60 and standard deviation of 12.5
```

```
set.seed(3)
```

```
sample <- rnorm(n, 60, 12.5)
```

```
# t.test function for sample data, confidence level, and selecting the $conf.int option
```

```
t.test(sample, conf.level = confidence_level)$conf.int
```

★+1 W3schools PathfinderTrack your progress - it's free!

[Log in](#)

[Sign Up](#)