

How can confidence intervals be calculated in Python?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can confidence intervals be calculated in Python?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136226>

Confidence intervals in Python can be calculated using various statistical libraries such as NumPy, SciPy, and Pandas. The process involves determining the mean and standard deviation of the data, choosing a confidence level, and using the appropriate formula to calculate the upper and lower bounds of the interval. These libraries also provide functions to graphically represent the confidence interval. Additionally, user-defined functions can be created using mathematical equations to calculate the confidence interval. Overall, confidence intervals can be easily calculated in Python by utilizing the available statistical tools and functions.

Calculate Confidence Intervals in Python

A confidence interval for a mean is a range of values that is likely to contain a population mean with a certain level of confidence.

It is calculated as:

$$\text{Confidence Interval} = \bar{x} \pm t^*(s/\sqrt{n})$$

where:

x: sample mean
t: t-value that corresponds to the confidence level
s: sample standard deviation
n: sample size

This tutorial explains how to calculate confidence intervals in Python.

Confidence Intervals Using the t Distribution

If we're working with a small sample ($n < 30$), we can use the `from the scipy.stats` library to calculate a confidence interval for a population mean.

The following example shows how to calculate a confidence interval for the true population mean height (in inches) of a certain species of plant, using a sample of 15 plants:

```
import numpy as np
import scipy.stats as st

#define sample data
data =

#create 95% confidence interval for population mean
weight
st.t.interval(alpha=0.95,          df=len(data)-1,
loc=np.mean(data), scale=st.sem(data))

(16.758, 24.042)
```

The 95% confidence interval for the true population mean height is (16.758, 24.042).

You'll notice that the larger the confidence level, the

wider the confidence interval. For example, here's how to calculate a 99% C.I. for the exact same data:

```
#create 99% confidence interval for same sample
st.t.interval(alpha=0.99,          df=len(data)-1,
loc=np.mean(data), scale=st.sem(data))

(15.348, 25.455)
```

The 99% confidence interval for the true population mean height is (15.348, 25.455). Notice that this interval is wider than the previous 95% confidence interval.

Confidence Intervals Using the Normal Distribution

If we're working with larger samples ($n \geq 30$), we can assume that the sampling distribution of the sample mean is normally distributed (thanks to the) and can instead use the from the `scipy.stats` library.

The following example shows how to calculate a confidence interval for the true population mean height (in inches) of a certain species of plant, using a sample of 50 plants:

```
import numpy as np
```

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

```
#define sample data
```

```
np.random.seed(0)
```

```
data = np.random.randint(10, 30, 50)
```

```
#create 95% confidence interval for population mean  
weight
```

```
st.norm.interval(alpha=0.95, loc=np.mean(data),  
scale=st.sem(data))
```

```
(17.40, 21.08)
```

The 95% confidence interval for the true population mean height is (17.40, 21.08).

And similar to the t distribution, larger confidence levels lead to wider confidence intervals. For example, here's how to calculate a 99% C.I. for the exact same data:

```
#create 99% confidence interval for same sample
```

```
st.norm.interval(alpha=0.99, loc=np.mean(data),  
scale=st.sem(data))
```

```
(16.82, 21.66)
```

The 95% confidence interval for the true population mean height is (17.82, 21.66).

How to Interpret Confidence Intervals

Suppose our 95% confidence interval for the true population mean height of a species of plant is:

95% confidence interval = (16.758, 24.042)

The way to interpret this confidence interval is as follows:

There is a 95% chance that the confidence interval of contains the true population mean height of plants.

Another way of saying the same thing is that there is only a 5% chance that the true population mean lies outside of the 95% confidence interval. That is, there's only a 5% chance that the true population mean height of plants is less than 16.758 inches or greater than 24.042 inches.