

How can I perform Grubbs' Test in Python?

Authored by
stats writer

April 18, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I perform Grubbs' Test in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136671>

Grubbs' Test is a statistical method used to detect and remove outliers from a dataset. In order to perform Grubbs' Test in Python, one can use the "grubbs" function from the "scipy.stats" module. This function takes in the dataset as an input and returns the outlier values as well as the critical value for the test. It uses the extreme studentized deviate (ESD) formula to calculate the critical value and compares it to the maximum value in the dataset. If the maximum value is greater than the critical value, it is considered an outlier and is removed from the dataset. This process is repeated until there are no more outliers detected. By using the "grubbs" function, one can easily and efficiently perform Grubbs' Test in Python to identify and remove outliers from their dataset.

Perform Grubbs' Test in Python

Grubbs' Test is used to identify the presence of outliers in a dataset. To use this test, a dataset should be approximately normally distributed and have at least 7 observations.

This tutorial explains how to perform Grubbs' Test in Python.

Grubbs' Test in Python

To perform Grubbs' Test in Python, we can use the `smirnov_grubbs()` function from the outlier_utils package, which uses the following syntax:

```
smirnov_grubbs.test(data, alpha=.05)
```

where:

data: A numeric vector of data values
alpha: The

significance level to use for the test. Default is .05

To use this function, you need to first install the outlier_utils package:

```
pip install outlier_utils
```

Once this package is installed, you can perform Grubbs' Test. The following examples illustrate how to do so.

Example 1: Two-Sided Grubbs' Test

The following code illustrates how to perform a two-sided Grubbs' test, which will detect outliers on both ends of the dataset.

```
import numpy as np  
from outliers import smirnov_grubbs as grubbs
```

```
#define data
```

```
data = np.array()
```

```
#perform Grubbs' test
```

```
grubbs.test(data, alpha=.05)
```

```
array()
```

This function simply returns an array with the outliers removed. In this case, the max value of 40 was an outlier, so it was removed.

Example 2: One-Sided Grubbs' Test

The following code illustrates how to perform a one-sided Grubbs' test for both the minimum value and the maximum value in a dataset:

```
import numpy as np
from outliers import smirnov_grubbs as grubbs

#define data
data = np.array()

#perform Grubbs' test to see if minimum value is an
outlier
grubbs.min_test(data, alpha=.05)

array()

#perform Grubbs' test to see if minimum value is an
outlier
grubbs.max_test(data, alpha=.05)

array()
```

Example 3: Extract the Index of the Outlier

The following code illustrates how to extract the index of the outlier value:

```
import numpy as np
from outliers import smirnov_grubbs as grubbs

#define data
data = np.array()

#perform Grubbs' test and identify index (if any) of the
outlier
grubbs.max_test_indices(data, alpha=.05)
```

This tells us that there is an outlier in index position 16 of the array.

Example 4: Extract the Value of the Outlier

The following code illustrates how to extract the value of the outlier:

```
import numpy as np
from outliers import smirnov_grubbs as grubbs
```

```
#define data
```

```
data = np.array()
```

```
#perform Grubbs' test and identify the actual value (if  
any) of the outlier
```

```
grubbs.max_test_outliers(data, alpha=.05)
```

This tells us that there is one outlier with a value of 40.

How to Handle Outliers

If Grubbs' Test identifies an outlier in your dataset, you have a few options:

- 1. Double check to make sure that the value is not a typo or a data entry error. Sometimes values that show up as outliers in datasets are simply typos made by an individual when entering the data. First, verify that the value was entered correctly before you make any further decisions.**

- 2. Assign a new value to the outlier. If the outlier turns out to be a result of a typo or data entry error, you may decide to assign a new value to it, such as the mean or the median of the dataset.**

3.Remove the outlier. If the value is a true outlier, you may choose to remove it if it will have a significant impact on your analysis.

ARABPSYCHOLOGY.COM