

What is Chauvenet's Criterion and can you provide an example?

Authored by
stats writer

April 19, 2024

RECOMMENDED CITATION

stats writer (2024). *What is Chauvenet's Criterion and can you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137046>

Chauvenet's Criterion is a statistical method used to identify and eliminate outliers in a dataset. It is based on the assumption that a data point that deviates significantly from the majority of the data is likely due to error or measurement bias. The criterion calculates the probability of a data point being considered an outlier and compares it to a predetermined threshold. If the probability is lower than the threshold, the data point is considered an outlier and can be removed from the dataset.

For example, in a study measuring the heights of individuals, a data point that is significantly lower or higher than the average height of the population may be identified as an outlier using Chauvenet's Criterion. This could be due to measurement error or misreporting of height. By removing this outlier, the dataset becomes more accurate and representative of the population.

Chauvenet's Criterion: Definition & Example

An outlier is an that lies abnormally far away from other values in a dataset. Outliers can be problematic because they can affect the results of an analysis.

One way to identify outliers in a dataset is to use Chauvenet's Criterion, which uses the following process:

1. For each individual value x_i in the dataset, calculate the deviation from the mean as:

$$\text{Deviation} = |x_i - \bar{x}| / s$$

where \bar{x} is the sample mean and s is the sample standard deviation.

2. Compare the deviations of each individual value to

the critical values of Chauvenet's Criterion Table below. For individual data values with deviations greater than those found in the table, declare those data values to be outliers.

ARABPSYCHOLOGY.COM

n	Critical Value
2	1.150
3	1.383
4	1.534
5	1.645
6	1.732
7	1.803
8	1.863
9	1.915
10	1.960
11	2.000
12	2.037
13	2.070
14	2.100
15	2.128
16	2.154
17	2.178
18	2.200
19	2.222
20	2.241
21	2.260
22	2.278
23	2.295
24	2.311
25	2.326
26	2.341
27	2.355
28	2.369
29	2.382
30	2.394
31	2.406
32	2.418
33	2.429
34	2.440
35	2.450
36	2.460
37	2.470
38	2.479
39	2.489
40	2.498
50	2.576
100	2.807
500	3.291
1000	3.481

Chauvenet's Criterion: An Example

Suppose we have the following dataset of 15 values:

Values
4
6
6
8
12
13
14
14
19
20
22
24
25
27
42

The sample mean for this dataset is $\bar{x} = 17.067$ and the sample standard deviation is $s = 10.096$. For each individual data value, we can calculate its deviation as:

$$\text{Deviation} = |x_i - \bar{x}| / s$$

For example:

The first data value would have a deviation of $|4 - 17.067| / 10.096 = 1.294$. The first data value would have a deviation of $|6 - 17.067| / 10.096 = 1.096$.

And so on.

We can use the same formula to calculate the deviation of each individual data value:

Values	Deviations
4	1.294
6	1.096
6	1.096
8	0.898
12	0.502
13	0.403
14	0.304
14	0.304
19	0.192
20	0.291
22	0.489
24	0.687
25	0.786
27	0.984
42	2.470

It turns out that the value 42 has a deviation greater than 2.128:

Values	Deviations
4	1.294
6	1.096
6	1.096
8	0.898
12	0.502
13	0.403
14	0.304
14	0.304
19	0.192
20	0.291
22	0.489
24	0.687
25	0.786
27	0.984
42	2.470

→ Greater than 2.128

Thus, the value 42 is the only outlier in this dataset.

Cautions on Using Chauvenet's Criterion

Chauvenet's Criterion makes the assumption that the values in a dataset are normally distributed. If this assumption is not met, then using Chauvenet's Criterion to identify outliers is likely not valid.

If you do use this method and find that a value is an outlier, you should first verify that the value is not a result of a data entry error. Sometimes data is simply entered incorrectly.

If the value is a true outlier, you may choose to remove it if it will have a significant impact on your overall analysis. Just be sure to mention that you removed an outlier when you report your results.

Also, this method should only be used on a given dataset once. For example, suppose we use this criterion to identify the value 42 as an outlier in the previous example and remove this value from the dataset.

We then shouldn't recalculate the mean and sample standard deviation and calculate the deviations once again to find more outliers.