

What is the Assumption of Normality in Statistics?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the Assumption of Normality in Statistics?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139931>

The Assumption of Normality in statistics refers to the assumption that the data being analyzed follows a normal distribution, also known as the bell curve. This means that the data is evenly distributed around the mean, with most values falling close to the mean and fewer values falling further away. This assumption is important in many statistical analyses as it allows for the use of certain mathematical models and procedures that rely on the normal distribution. It also enables researchers to make more accurate predictions and inferences based on the data. However, it is important to note that this assumption may not always hold true in real-world data, and there are other methods available for analyzing non-normally distributed data.

What is the Assumption of Normality in Statistics?

Many statistical tests rely on something called the assumption of normality.

This assumption states that if we collect many independent random samples from a population and calculate some value of interest (like the) and then create a histogram to visualize the distribution of sample means, we should observe a perfect .

Many statistical techniques make this assumption about the data, including:

- 1. : It's assumed that the sample data is normally distributed.**
- 2. : It's assumed that both samples are normally distributed.**

3. : It's assumed that the residuals from the model are normally distributed.

4. : It's assumed that the residuals from the model are normally distributed.

If this assumption is violated then the results of these tests become unreliable and we're unable to generalize our findings from the sample data to the overall with confidence. This is why it's import to check if this assumption is met.

There are two common ways to check if this assumption of normality is met:

1. Visualize Normality

2. Perform a Formal Statistical Test

The following sections explain the specific graphs you can create and the specific statistical tests you can perform to check for normality.

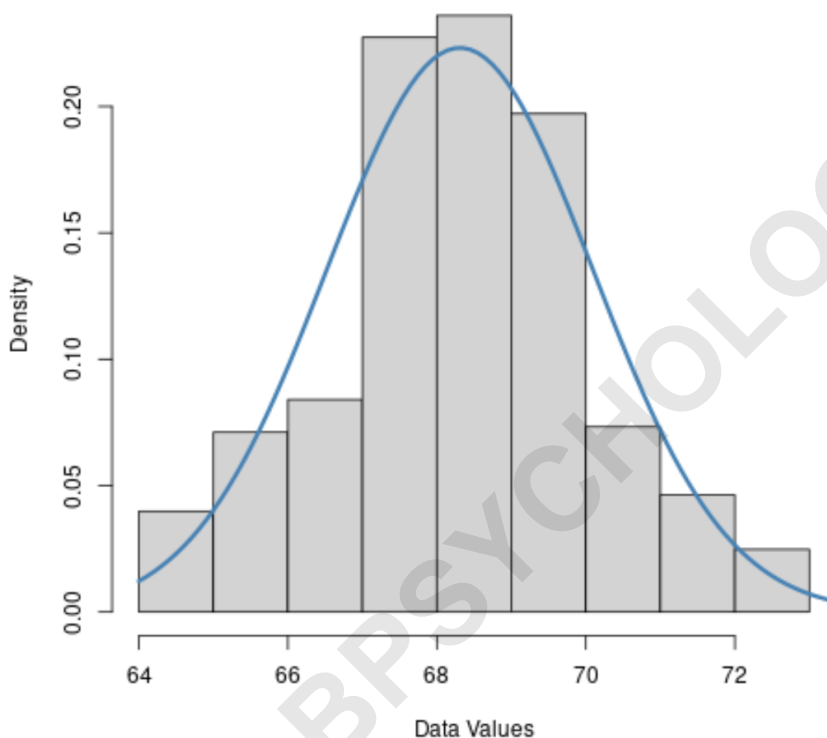
Visualize Normality

A quick and informal way to check if a dataset is normally distributed is to create a histogram or a Q-Q

plot.

1. Histogram

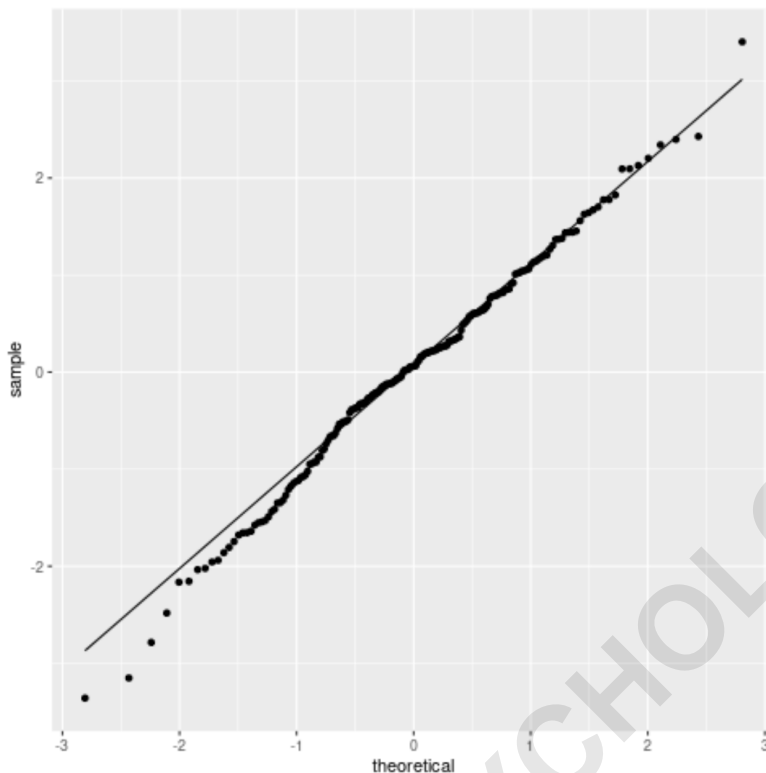
If a histogram for a dataset is roughly bell-shaped, then it's likely that the data is normally distributed.



A Q-Q plot, short for "quantile-quantile" plot, is a type of plot that displays theoretical quantiles along the x-axis (i.e. where your data would lie if it did follow a normal distribution) and sample quantiles along the y-axis (i.e. where your data actually lies).

If the data values fall along a roughly straight line at a

45-degree angle, then the data is assumed to be normally distributed.



Perform a Formal Statistical Test

You can also perform a formal statistical test to determine if a dataset is normally distributed.

If the of the test is less than a certain significance level (like $\alpha = 0.05$) then you have sufficient evidence to say that the data is *not* normally distributed.

There are three statistical tests that are commonly used to test for normality:

1. The Jarque-Bera Test

2. The Shapiro-Wilk Test

3. The Kolmogorov-Smirnov Test

How to Perform a Kolmogorov-Smirnov Test in Excel

What to Do if the Assumption of Normality is Violated

If it turns out that your data is not normally distributed then you have two options:

1. Transform the data.

One option is to simply *transform* the data to make it more normally distributed. Common transformations include:

Log Transformation: Transform the data from y to $\log(y)$.

Square Root Transformation: Transform the data

from y to \sqrt{y}

Cube Root Transformation: Transform the

data from y to $y^{1/3}$

Box-Cox Transformation: Transform

the data using a

By performing these transformations, the distribution of data values typically becomes more normally

distributed.

2. Perform a Non-Parametric Test

Statistical tests that make the assumption of normality are known as *parametric tests*. But there are also a family of tests known as *non-parametric tests* that do not make this assumption of normality.

If it turns out that your data is not normally distributed, you could simply perform a non-parametric test. Here are a few non-parametric versions of common statistical tests:

Parametric Test	Non-Parametric Equivalent
One Sample t-test	One Sample Wilcoxon Signed Rank Test
Two Sample t-test	
Paired Samples t-test	
One-Way ANOVA	

Each of these non-parametric tests allow you to perform a statistical test without satisfying the assumption of normality.