

Statistics. Simplified.

A grey icon consisting of three horizontal lines of varying lengths, positioned to the left of the word "MENU".

MARCH 17, 2021 BY ZACH

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 **sample mean**) and then create a histogram to visualize the distribution of sample means, we should observe a perfect **bell curve**.

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

1. **One sample t-test:** It's assumed that the sample data is normally distributed.
2. **Two sample t-test:** It's assumed that both samples are normally distributed.
3. **ANOVA:** It's assumed that the residuals from the model are normally distributed.

4. Linear regression: 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 **population** with confidence. This is why it's important 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

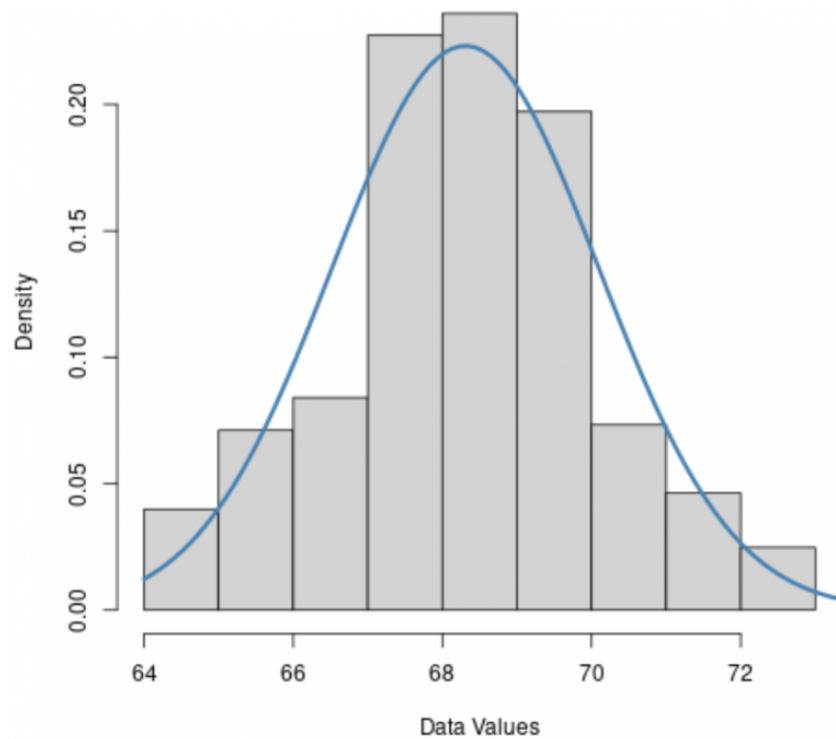
The following sections explain the specific graphs you can create and the specific statistical tests you can perform to check for normality.

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

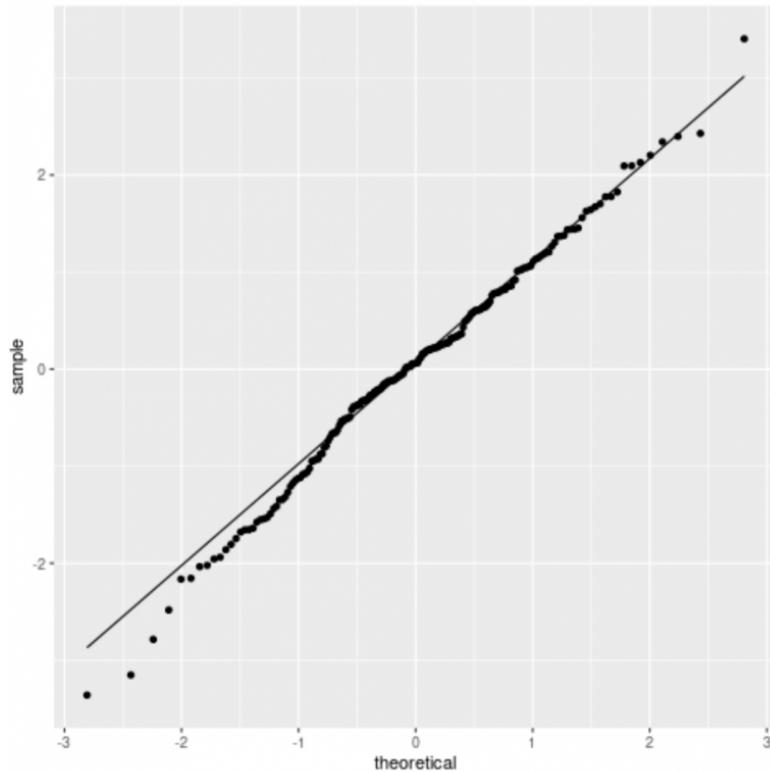
If a histogram for a dataset is roughly bell-shaped, then it's likely that the data is normally distributed.



2. Q-Q Plot

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 **p-value** 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

- How to Perform a Jarque-Bera Test in Excel
- How to Perform a Jarque-Bera Test in R
- How to Perform a Jarque-Bera Test in Python

2. The Shapiro-Wilk Test

- How to Perform a Shapiro-Wilk Test in R
- How to Perform a Shapiro-Wilk Test in Python

3. The Kolmogorov-Smirnov Test

- How to Perform a Kolmogorov-Smirnov Test in Excel
- How to Perform a Kolmogorov-Smirnov Test in R
- How to Perform a Kolmogorov-Smirnov Test in Python

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 Box-Cox procedure

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	Mann-Whitney U Test
Paired Samples t-test	Two Sample Wilcoxon Signed Rank Test
One-Way ANOVA	Kruskal-Wallis Test

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

Additional Resources

[The Four Assumptions Made in a T-Test](#)

[The Four Assumptions of Linear Regression](#)

[The Four Assumptions of ANOVA](#)



Published by Zach

[View all posts by Zach](#)

PREV

[The Four Assumptions Made in a T-Test](#)

NEXT

[Hedges' g Calculator](#)