Normality Tests and What They Can Tell You

Pick a fixed example sample below to see how a Q-Q plot pattern, a Shapiro-Wilk result, and an Anderson-Darling result each describe the same data.

StatReason

Normality Tests and What They Can Tell You

Example sample

No detectable departure from normality at this sample size.

Q-Q plot pattern	Points closely follow the diagonal line across the full range.
Shapiro-Wilk result	W = 0.991, p = .74 (not significant)
Anderson-Darling result	A^2 = 0.28, p = .63 (not significant)
What this means	No detectable departure from normality at this sample size.

Run these tests on your own data: **Shapiro-Wilk Test**

What this answers

This page answers "which normality check should I trust, and what does a pass or fail actually tell me?" A Q-Q plot, Shapiro-Wilk test, and Anderson-Darling test all check the same underlying question, approximate normality, but they respond differently to sample size and to specific kinds of departure from normality.

Three tools, one question

A Q-Q plot compares your sorted data against the values a normal distribution would predict at each rank; points that fall close to a straight diagonal line suggest an approximately normal shape, while systematic curvature or a bend at one end suggests skew or heavy tails. Shapiro-Wilk and Anderson-Darling are formal tests that produce a p-value against the null model of normality, but they answer "is there detectable evidence against normality," not "is this distribution exactly normal."

Sample size changes what a test can detect

With a small sample, even a Q-Q plot with visible bends may not produce a statistically significant Shapiro-Wilk or Anderson-Darling result, because there is not enough data to distinguish random scatter from a genuine departure. With a very large sample, both tests can flag a statistically significant departure from normality that is too small to matter for the method you plan to use, since most standard tests are fairly robust to mild non-normality at large sample sizes.

Worked example

Select "Right-skewed, n = 100" above. The Q-Q plot pattern bends noticeably away from the diagonal in the upper tail, and both formal tests return a statistically significant result at this sample size, agreeing with the plot. Select "Normal-like, n = 12" and both formal tests are likely to return a nonsignificant result even if the sample looks slightly uneven by eye, because 12 observations rarely provide enough evidence either way.

Assumption audit

> **Calculated from your data:** the actual Shapiro-Wilk or Anderson-Darling statistic and p-value, and the exact Q-Q plot coordinates, once you run your own data through the linked engines.

> **Evidence to review:** whether the departure shown, if any, is concentrated in the tails, in one direction (skew), or spread evenly, since these patterns call for different responses.

> **You must verify:** whether the method you plan to use is sensitive to non-normality at your actual sample size; many common tests remain valid under mild departures, especially with larger samples.

Source

This comparison follows the normality-assessment guidance in the NIST/SEMATECH e-Handbook of Statistical Methods and the shared statistical reasoning contract every StatReason engine is built against.

Limitations

No normality check, visual or formal, can prove a distribution is exactly normal; all of them can only fail to detect a departure or detect one that may or may not matter for your specific method.