

# Anderson–Darling test

In [statistics](#), the **Anderson–Darling test**, named after [Theodore Wilbur Anderson](#) (born 1918) and [Donald A. Darling](#) (born 1915), who invented it in 1952<sup>[1]</sup>, is a [statistical test](#) of whether there is evidence that a given sample of data did not arise from a given [probability distribution](#). In its basic form, the test assumes that there are no parameters to be estimated in the distribution being tested, in which case the test and its set of [critical values](#) is distribution-free. However, the test is most often used in contexts where a family of distributions is being tested, in which case the parameters of that family need to be estimated and account must be taken of this in adjusting either the test-statistic or its critical values.

When applied to testing if a [normal distribution](#) adequately describes a set of data, it is one of the most powerful statistical tools for detecting most departures from [normality](#).<sup>[2][3]</sup>

In addition to its use as a test of fit for distributions, it can be used in parameter estimation as the basis for a form of [minimum distance estimation](#) procedure.

**K-sample Anderson–Darling tests** are available for testing whether several collections of observations can be modelled as coming from a single population, where the [distribution function](#) does not have to be specified.

## The single-sample test

The Anderson–Darling and [Cramér–von Mises statistics](#) belong to the class of quadratic [EDF](#) statistics (tests based on the [empirical distribution function](#)).<sup>[3]</sup> If the hypothesized distribution is  $F$ , and empirical (sample) cumulative distribution function is  $F_n$ , then the quadratic EDF statistics measure the distance between  $F$  and  $F_n$  by

$$n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 w(x) dF(x),$$

where  $w(x)$  is a weighting function. When the weighting function is  $w(x)=1$ , the statistic is the [Cramér–von Mises statistic](#). The Anderson–Darling (1954) test<sup>[4]</sup> is based on the distance

$$n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{[F(x) (1 - F(x))]} dF(x),$$

which is obtained when the weight function is  $w(x) = [F(x) (1 - F(x))]^{-1}$ . Thus, compared with the [Cramér–von Mises distance](#), the Anderson–Darling distance places more weight on observations in the tails of the distribution.

## Basic test statistic

The Anderson–Darling test assesses whether a [sample](#) comes from a specified distribution. It makes use of the fact that, when given a hypothesized underlying distribution and assuming the data does arise from this distribution, the data can be transformed to a [uniform distribution](#). The transformed sample data can be then tested for uniformity with a distance test (Shapiro 1980). The formula for the [test statistic](#)  $A$  to assess if data  $\{Y_1 < \dots < Y_n\}$  (note that the data must be put in order) comes from a distribution with [cumulative distribution function](#) (CDF)  $F$  is

$$A^2 = -n - S,$$

where

$$S = \sum_{k=1}^n \frac{2k-1}{n} [\ln F(Y_k) + \ln (1 - F(Y_{n+1-k}))].$$

The test statistic can then be compared against the critical values of the theoretical distribution. Note that in this case no parameters are estimated in relation to the distribution function  $F$ .

Essentially the same test statistic can be used in the test of fit of a family of distributions, but then it must be compared against the critical values appropriate to that family of theoretical distributions and dependent also on the method used for parameter estimation.

## Test for normality

In comparisons of power, Stephens<sup>[2]</sup> found  $A^2$  to be one of the best [Empirical distribution function](#) statistics for detecting most departures from normality. The only statistic close was the  $W^2$  [Cramér–von Mises test](#) statistic. It may be used with small sample sizes  $n \leq 25$ . Very large sample sizes may reject the assumption of normality with only slight imperfections, but industrial data with sample sizes of 200 and more have passed the Anderson–Darling test.<sup>[[citation needed](#)]</sup>

The computation differs based on what is known about the distribution:<sup>[5]</sup>

- Case 1: The mean  $\mu$  and the variance  $\sigma^2$  are both known.
- Case 2: The variance  $\sigma^2$  is known, but the mean  $\mu$  is unknown.
- Case 3: The mean  $\mu$  is known, but the variance  $\sigma^2$  is unknown.
- Case 4: Both the mean  $\mu$  and the variance  $\sigma^2$  are unknown.

The  $n$  observations  $X_i$ , for  $i = 1, \dots, n$ , of the variable  $X$  that should be tested are sorted from low to high and the notation in the following assumes that  $X_i$  represent the ordered observations. Let

$$\hat{\mu} = \begin{cases} \mu & , \text{ if the mean is known.} \\ \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i & , \text{ otherwise.} \end{cases}$$

$$\hat{\sigma}^2 = \begin{cases} \sigma^2 & , \text{ if the variance is known.} \\ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 & , \text{ if the variance is not known, but the mean is.} \\ \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 & , \text{ otherwise.} \end{cases}$$

The values  $X_i$  are standardized to create new values  $Y_i$ , given by

$$Y_i = \frac{X_i - \hat{\mu}}{\hat{\sigma}}.$$

With the standard normal CDF  $\Phi$ ,  $A^2$  is calculated using

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) (\ln \Phi(Y_i) + \ln(1 - \Phi(Y_{n+1-i}))).$$

An alternative expression in which only a single observation is dealt with at each step of the summation is:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1) \ln \Phi(Y_i) - (2(n - i) + 1) \ln(1 - \Phi(Y_i))].$$

A modified statistic is calculated using

$$A^{*2} = \begin{cases} A^2 \left(1 + \frac{4}{n} - \frac{25}{n^2}\right) & , \text{ if the variance and the mean are both unknown.} \\ A^2 & , \text{ otherwise.} \end{cases}$$

If  $A^{*2}$  exceeds a given critical value, then the hypothesis of normality is rejected with some significance level. The critical values are given in the table below (valid for  $n \geq 5$ ).<sup>[2]</sup>

Note 1: If  $\hat{\sigma} = 0$  or any  $\Phi(Y_i) = (0 \text{ or } 1)$  then  $A^2$  cannot be calculated and is undefined.

Note 2: The above adjustment formula is taken from Shorak & Wellner (1986, p239). Care is required in comparisons across different sources as often the specific adjustment formula is not stated.

Note 3: Stephens<sup>[2]</sup> notes that the test becomes better when the parameters are computed from the data, even if they are known.

**Significance Case 1 Case 2\* Case 3 Case 4**

15%          1.610   ?          ?          0.576

10%	1.933	0.908	1.760	0.656
5%	2.492	1.105	2.323	0.787
2.5%	3.070	1.304	2.904	0.918
1%	3.857	1.573	3.690	1.092

(\*) For the case 2, the values are for the asymptotic distribution.

Alternatively, for case 3, D'Agostino (1986, pp. 372-373)<sup>[5]</sup> gives the adjusted statistic

$$A^{*2} = A^2 \left( 1 + \frac{0.75}{n} + \frac{2.25}{n^2} \right).$$

and normality is rejected if  $A^{*2}$  exceeds 0.631, 0.752, 0.873, 1.035, or 1.159 at 10%, 5%, 2.5%, 1%, and 0.5% significance levels, respectively; the procedure is valid for sample size at least  $n=8$ .

## Tests for other distributions

Above, it was assumed that the variable  $X_i$  was being tested for normal distribution. Any other family of distributions can be tested but the test for each family is implemented by using a different modification of the basic test statistic and this is referred to critical values specific to that family of distributions. The modifications of the statistic and tables of critical values are given by Stephens (1986)<sup>[3]</sup> for the exponential, extreme-value, Weibull, gamma, logistic, Cauchy, and von Mises distributions. Tests for the (two-parameter) [log-normal distribution](#) can be implemented by transforming the data using a logarithm and using the above test for normality. Details for the required modifications to the test statistic and for the critical values for the [normal distribution](#) and the [exponential distribution](#) have been published by Pearson & Hartley (1972, Table 54). Details for these distributions, with the addition of the [Gumbel distribution](#), are also given by Shorak & Wellner (1986, p239). Details for the [logistic distribution](#) are given by Stephens (1979). A test for the (two parameter) [Weibull distribution](#) can be obtained by making use of the fact that the logarithm of a Weibull variate has a [Gumbel distribution](#).

## Non-parametric k-sample tests

Scholz F.W. and Stephens M.A. (1987) discuss a test, based on the Anderson-Darling measure of agreement between distributions, for whether a number of random samples with possibly different sample sizes may have arisen from the same distribution, where this distribution is unspecified.