Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.
https://simfit.org.uk
https://simfit.silverfrost.com

The Kolmogorov-Smirnov nonparametric test is used to check if a sample is consistent with a known continuous distribution. It is most powerful with large samples from such a distribution when the true parameters are known, and is much weaker if parameters have to be estimated from the sample, or if a discontinuous distribution is to be considered.

To be precise, the user has a sample (i.e. vector X) of n observations

$$X = (x_1, x_2, \ldots, x_n)$$

and wishes to test if these numbers are consistent with a known distribution, where the parameters have been previously estimated with great precision from an independent very large sample, or are known due to further information. Preferably the data should cover a wide range and n should not be too small, say $n > 20$?

The test is based upon the largest vertical distance where the sample cumulative exceeds the theoretical distribution ($D+$), the largest vertical distance where the theoretical distribution exceeds the sample cumulative distribution ($D-$), or the the maximum of these ($D = \max(D+, D-)$). The standardized Z values are defined as $Z = D\sqrt{n}$, and SIMFIT calculates exact p for small samples, but uses a series expansion for Z with large samples.

Choose `[A/Z]` from the main SIMFIT menu, open program **simstat**, select statistical tests, then choose the 1-sample Kolmogorov-Smirnov test. First of all select to test for a uniform distribution $U(A, B)$ with $A = 0$ and $B = 2$ to get the following results.

Kolmogorov-Smirnov one sample test 1: Uniform(A,B)

Data: test file g08cbf.tf1 (Kolmogorov-Smirnov 1-sample test)
Parameters fixed by user: A = 0, B = 2
Sample size = 30, i.e. number of x-values

H_0: $F(x)$ equals $G(y)$ (x and theory are comparable) against
H_1: $F(x)$ not equal to $G(y)$ (x and theory not comparable)

D	0.2800	
Z	1.534	
p	0.0143	Reject H_0 at 5% significance level

H_2: $F(x) > G(y)$ (x tend to be smaller than theoretical)

D	0.2800	
Z	1.534	
p	0.0071	Reject H_0 at 1% significance level

H_3: $F(x) < G(y)$ (x tend to be larger than theoretical)

D	0.02333
Z	0.1278
p	0.5000

Here $F(x)$ is the sample distribution while $G(y)$ is the theoretical distribution, and these figures are interpreted as follows. The three D values were

$$D+ = 0.28$$
$$D- = 0.02333$$
$$D = 0.28$$

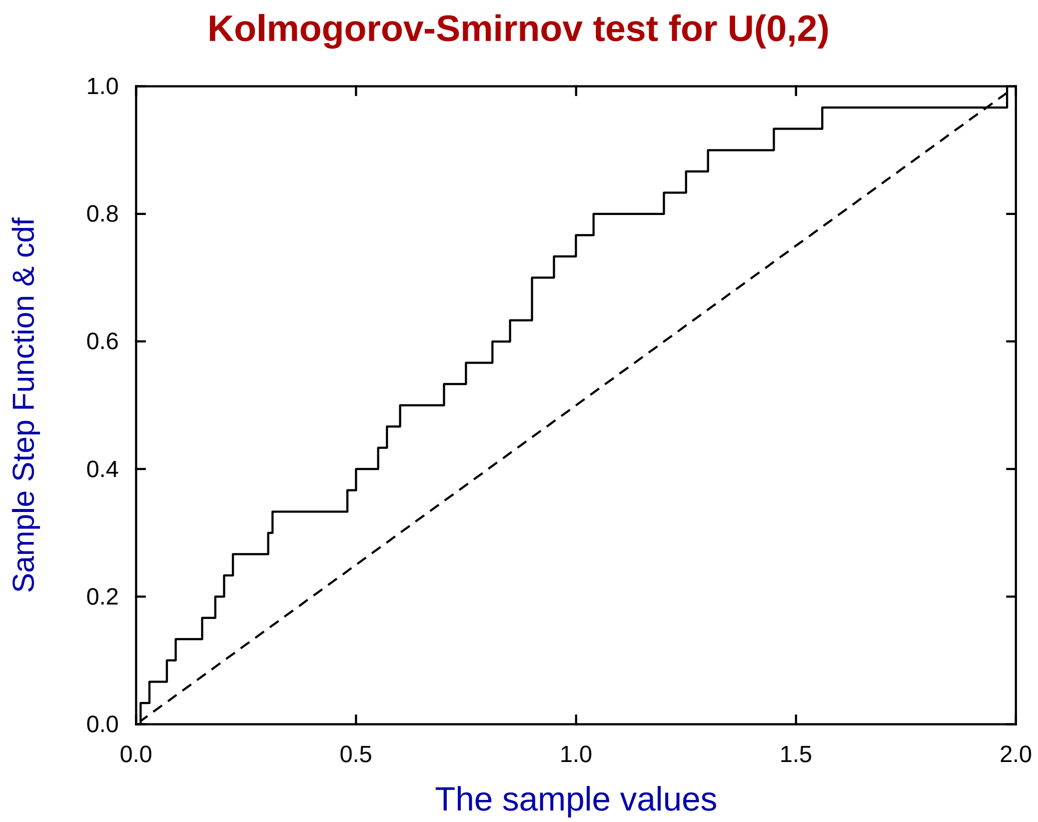
and the three cases are therefore as follows.

1. Reject H_0 against H_1 as a two-tail test indicates $F(x)$ and $G(y)$ are unlikely to be equal ($p = 0.0143$).

2. Reject H_0 against H_2 as a one-tail test indicates that $F(x)$ is more likely to be larger than $G(y)$ than to be equal to it ($p = 0.0071$).

3. Do not reject H_0 against H_3 as a one tail test offers no support for the case that $F(x)$ is smaller than $G(y)$ ($p = 0.5$).

These results clearly reject the case $F(x) = G(y)$, in favor of H_1, i.e. that x values tend to be smaller than y values, indicating that the sample cumulative distribution was heavily displaced to the left of the theoretical distribution. To confirm this interpretation visually we can plot the sample cumulative distribution and theoretical distribution as in the next graph.



Having rejected the null hypothesis H_0: the sample is distributed as $U(0,2)$, we can try another theoretical distribution.

So, in this next case, we proceed to test the null hypothesis that the theoretical distribution is normal with parameters $\mu = 0.6967$ and $\sigma^2 = 0.2564$ as estimated from the sample. We conclude that this cannot be rejected and that the sample distribution is close to the theoretical distribution, as displayed graphically.

Kolmogorov-Smirnov one sample test 2: Normal(μ, σ^2)

Parameters estimated from sample are:
μ = 6.967E-01, se = 9.244E-02, 95%confidence limits = (5.076E-01, 8.857E-01)
σ = 5.063E-01, σ^2 = 2.564E-01, 95%cl = (1.626E-01,4.633E-01)

H_0: $F(x)$ equals $G(y)$ (x and theory are comparable) against
H_1: $F(x)$ not equal to $G(y)$ (x and theory not comparable)

D	0.1108
Z	0.6068
p	0.8162

H_2: $F(x) > G(y)$ (x tend to be smaller than theoretical)

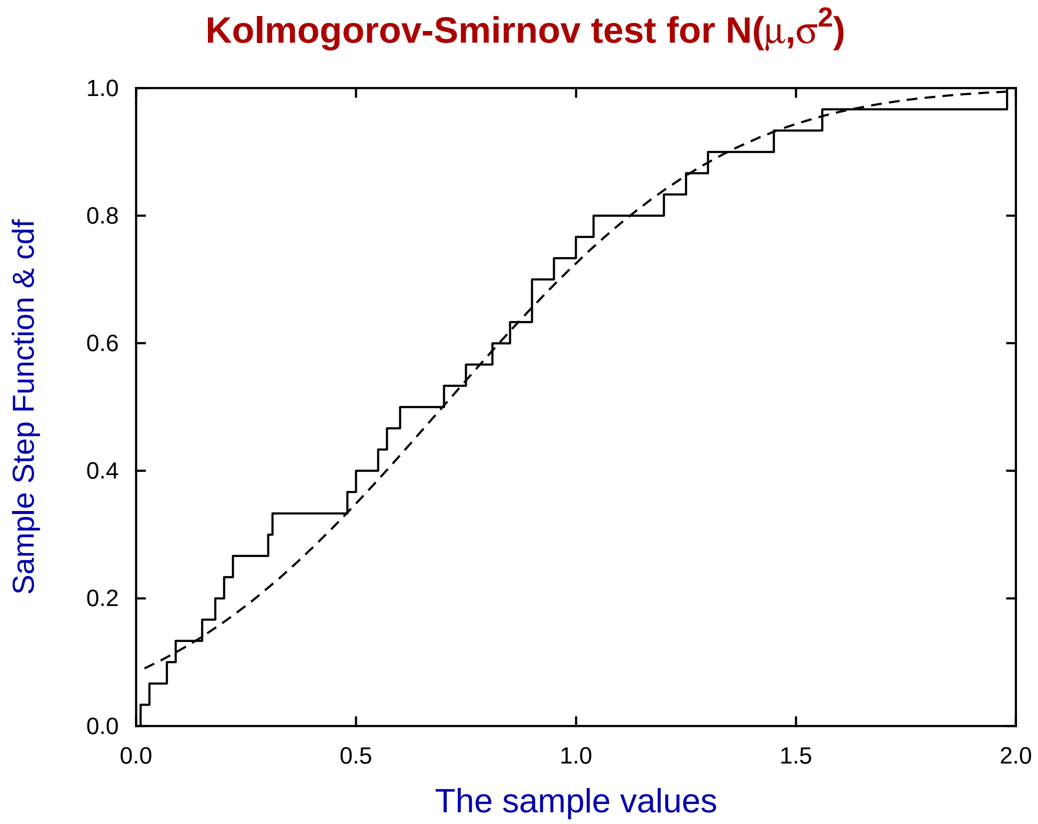
D	0.1108
Z	0.6068
p	0.4081

H_3: $F(x) < G(y)$ (x tend to be larger than theoretical)

D	0.08753
Z	0.4794
p	0.4801

Shapiro-Wilks normality test (Note: Bonferroni $n \geq 2$):

W	0.9529	
p	0.2019	Tentatively accept normality



Note that SIMFIT often presents the results from several tests at the same time, which is convenient for preliminary data investigation but not for precise analysis. Which is why, as in the last table, hints about the Bonferroni principle are frequently given.