



Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.
<http://www.simfit.org.uk>

The Kolmogorov-Smirnov two sample nonparametric test can be used to examine if it is reasonable to conclude that two sets of independent observations have the same unknown distribution. It is based on a test statistic estimated from the largest differences between the two sample cumulative distributions.

To be precise, the user has two samples (i.e. vectors X and Y) with m and n observations

$$X = (x_1, x_2, \dots, x_m)$$

$$Y = (y_1, y_2, \dots, y_n)$$

and wishes to test the null hypothesis that the samples have the same distribution, against the alternative hypothesis that they are not equal, or possibly the one-sided alternatives. That is

$$H_0 : F(x) = G(y)$$

$$H_1 : F(x) \neq G(y)$$

$$H_2 : F(x) > G(y)$$

$$H_3 : F(x) < G(y)$$

and SIMFIT provides all the information that is required to perform such tests.

From the main SIMFIT menu select [A/Z], choose to open program **rstest**, then analyze the test files provided to obtain the following results.

Kolmogorov-Smirnov two sample test 1

X-data: Test file ttest.tf2: for paired t test with TTEST.TF3

Y-data: Test file ttest.tf3: for paired t test with TTEST.TF2

X sample size 10

Y sample size 10

H_0 : $F(x)$ is equal to $G(y)$ (x and y are comparable) against

H_1 : $F(x)$ not equal to $G(y)$ (x and y are not comparable)

D 0.2000

Z 0.08944

p 0.7869

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

D 0.2000

Z 0.08944

p 0.3935

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

D 0.1000

Z 0.04472

p 0.4972

The test statistic uses D , which would be either the largest difference where $F(x) > G(y)$, D_m , the largest difference where $F(x) < G(y)$, D_n , or the maximum of these, $D_{mn} = \max(D_m, D_n)$, and is defined as

$$Z = \sqrt{\frac{mn}{m+n}} D.$$

This test is very weak unless the distributions are continuous and the sample sizes fairly large, so it is not surprising that the first example does not lead to a rejection of the null hypothesis. However, the next example compares data from two test files, `normal.tf1` with 50 pseudo random numbers from a normal distribution with $\mu = 0$ and $\sigma = 1$, and `normal.tf2` with 50 pseudo random numbers from a normal distribution with $\mu = 1$ and $\sigma = 2$. The plot confirms very convincingly the table results, that identical distribution can be rejected in favor of the alternative hypothesis that $X < Y$, but not in favor of the alternative that $X > Y$.

Kolmogorov-Smirnov two sample test 2

X-data: Test file normal.tf1: mean = 0, standard deviation = 1
Y-data: Test file normal.tf2: mean = 1, standard deviation = 2
X sample size 50
Y sample size 50
 H_0 : $F(x)$ is equal to $G(y)$ (x and y are comparable) against
 H_1 : $F(x)$ not equal to $G(y)$ (x and y are not comparable)
D 0.3600
Z 0.07200
p 0.0013 Reject H_0 at 1% significance level
 H_2 : $F(x) > G(y)$ (x tend to be smaller than y)
D 0.3600
Z 0.07200
p 0.0007 Reject H_0 at 1% significance level
 H_3 : $F(x) < G(y)$ (x tend to be larger than y)
D 0.06000
Z 0.01200
p 0.4989

Kolmogorov-Smirnov Two Sample Test

