

Understanding and Performing the Kolmogorov-Smirnov Test in Excel

Authored by
Mohammed loot

November 1, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Understanding and Performing the Kolmogorov-Smirnov Test in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=7735>

Understanding the Kolmogorov-Smirnov Test Fundamentals

The [Kolmogorov-Smirnov test](#) (often abbreviated as the K-S test) stands as a foundational and indispensable tool in statistical analysis. It is classified as a [non-parametric statistical procedure](#) used primarily to assess whether a particular sample of observations plausibly originated from a theoretical distribution. This specific application is known as a goodness-of-fit test, and the most common theoretical target distribution is the [Normal Distribution](#).

Grabbing the distributional characteristics of data is paramount before proceeding with advanced modeling. Many powerful [statistical tests](#), including t-tests, linear regression, and Analysis of Variance (ANOVA), rely on the strict assumption that the underlying data population is normally distributed. If this core assumption is violated, the results derived from these parametric tests may be fundamentally flawed, leading to inaccurate conclusions and invalid inferences.

The K-S test provides a quantitative measure of fit by directly comparing the observed data's distribution against the expected theoretical distribution. It achieves this by calculating the maximum vertical distance between the sample's [Empirical Cumulative Distribution Function](#) (ECDF) and the Cumulative Distribution Function (CDF) of the hypothesized theoretical model. This detailed guide walks through the precise methodology for executing this critical analysis using standard functions readily available within Microsoft Excel.

The Critical Role of Normality Verification

A sound data analysis workflow must always begin with robust verification of underlying assumptions. Data analysts are obligated to confirm that their datasets satisfy all preconditions prior to employing any parametric modeling technique, with the assumption of normality being the most frequently tested requirement. Skipping this crucial normality check jeopardizes the reliability and validity of all subsequent hypothesis testing and predictive modeling efforts.

While preliminary visual diagnostics, such as generating histograms to inspect symmetry or creating Q-Q plots to evaluate quantile alignment, can offer intuitive insights into the data's shape and kurtosis, they often lack the objective rigor necessary for formal statistical decisions. The K-S test overcomes this subjectivity by yielding a definitive test statistic (D). This single value permits a formal, objective decision based on established critical values, moving the analysis beyond mere visual inspection.

The following technical methodology outlines the precise, step-by-step calculations required to compute the K-S statistic (D) entirely within the Excel environment. By following these steps, practitioners ensure that their assessment of distributional assumptions adheres strictly to standard statistical principles, thereby validating the foundation for any subsequent parametric analysis.

Step 1: Data Entry and Preliminary Sorting

The initial phase of conducting the Kolmogorov-Smirnov test demands meticulous preparation and structuring of the raw dataset. For instructional clarity, we will work with a modest sample dataset consisting of $n = 20$ observations. It is absolutely essential that all data values are entered into a single column (e.g., Column A) and subsequently sorted in **ascending order**. The subsequent calculation of the empirical cumulative distribution relies entirely upon the accurate numerical ranking established by this sorting process.

To begin the process, the numerical values for the dataset should be entered as shown below, ensuring they are arranged from the smallest value to the largest:

	A	B	C	D	E	F
1	Data					
2	5.57					
3	8.32					
4	8.35					
5	8.74					
6	8.75					
7	9.38					
8	9.91					
9	9.96					
10	10.36					
11	10.65					
12	10.77					
13	10.97					
14	11.15					
15	11.18					
16	11.47					
17	11.64					
18	11.88					
19	12.24					
20	13.02					
21	13.19					
22						
23						
24						
25						

Establishing this organized, sorted structure is non-negotiable, as it guarantees that we can accurately compute the rank and corresponding cumulative probabilities required for a precise comparison against the theoretical probabilities derived from the hypothesized normal distribution in the following analytical steps.

Step 2: Calculating Empirical and Theoretical Probabilities

This core analytical phase involves two primary sub-tasks: first, calculating the essential sample statistics (the mean and standard deviation) to parameterize the theoretical model; and second, using these parameters to derive the empirical and theoretical cumulative probabilities for every data point. The ultimate objective is to pinpoint the maximum absolute deviation between these two distributions, which yields the final K-S test statistic, D .

The computation necessitates the creation of several intermediate columns to track the ranks, cumulative probabilities, and differences, as illustrated in the visual aid below:

	A	B	C	D	E	F	G	H	I	J
1	Data	Cumulative	Expected	(RANK-1)/N	NORM.S.INV	Actual	Difference		Mean	10.375
2	5.57	1	0.05	0	-1.64485363	0.00426	0.004264		Std. Dev.	1.82672
3	8.32	2	0.1	0.05	-1.28155157	0.1303	0.080302			
4	8.35	3	0.15	0.1	-1.03643339	0.13381	0.033814		Maximum	0.10983
5	8.74	4	0.2	0.15	-0.84162123	0.18538	0.035381			
6	8.75	5	0.25	0.2	-0.67448975	0.18685	0.013152			
7	9.38	6	0.3	0.25	-0.52440051	0.29298	0.042983			
8	9.91	7	0.35	0.3	-0.38532047	0.39953	0.099534			
9	9.96	8	0.4	0.35	-0.2533471	0.41014	0.060141			
10	10.36	9	0.45	0.4	-0.12566135	0.49672	0.096724			
11	10.65	10	0.5	0.45	0	0.55983	0.109832			
12	10.77	11	0.55	0.5	0.125661347	0.5856	0.085597			
13	10.97	12	0.6	0.55	0.253347103	0.62768	0.077682			
14	11.15	13	0.65	0.6	0.385320466	0.66431	0.064311			
15	11.18	14	0.7	0.65	0.524400513	0.67028	0.020278			
16	11.47	15	0.75	0.7	0.67448975	0.72556	0.025558			
17	11.64	16	0.8	0.75	0.841621234	0.75569	0.005688			
18	11.88	17	0.85	0.8	1.036433389	0.795	0.005004			
19	12.24	18	0.9	0.85	1.281551566	0.84636	0.003638			
20	13.02	19	0.95	0.9	1.644853627	0.92618	0.026184			
21	13.19	20	1	0.95		0.93834	0.011657			
22										
23										
24										
25										

The successful implementation of the K-S test in Excel hinges upon the accurate calculation of specific summary statistics and cumulative probabilities. The following formulas are entered into Row 2 and must be copied down to encompass all $n=20$ observations (Row 21):

B2 (Rank): =ROW() - 1. This simple function assigns the numerical rank to the observation based on its position within the sorted list.

C2 (Empirical CDF P1): =B2/COUNT(\$A\$2:\$A\$21). This formula calculates the upper bound of the

empirical cumulative probability, $P(X \leq x_i)$, representing the proportion of values less than or equal to the current observation.

D2 (Empirical CDF P2): $= (B2-1) / \text{COUNT}(\$A\$2:\$A\$21)$. This calculates the lower bound of the empirical cumulative probability, $P(X < x_i)$, which is crucial for defining the interval and calculating the difference against the theoretical **CDF**.

E2 (Standard Normal Quantile): $= \text{IF}(C2 < 1, \text{NORM.S.INV}(C2), " ")$. This function translates the empirical probability (P1) into the corresponding **Z-score**, or quantile value, assuming the characteristics of a standard normal distribution (mean=0, standard deviation=1).

F2 (Theoretical CDF): $= \text{NORM.DIST}(A2, \$J\$1, \$J\$2, \text{TRUE})$. This calculates the probability of observing a value less than or equal to the current observation (A2), based on the parameters of the hypothesized **Normal Distribution** defined by the sample Mean (μ) and Standard Deviation (σ). The final **TRUE** argument specifies that the function should return the cumulative probability.

G2 (Absolute Difference D): $= \text{ABS}(F2-D2)$. This step calculates one of the two necessary absolute differences: the vertical distance between the theoretical CDF and the lower bound of the empirical CDF. The K-S statistic requires finding the maximum of two differences (one using P1 and one using P2), but for this simplified approach, we focus on the calculation that typically reveals the maximum deviation.

For the theoretical CDF calculation (Column F) to be statistically sound, the following summary statistics must be accurately computed first, typically placed in a dedicated summary section of the spreadsheet:

J1 (Sample Mean, μ): $= \text{AVERAGE}(A2:A21)$.

J2 (Sample Standard Deviation, σ): $= \text{STDEV.S}(A2:A21)$. We specifically utilize the sample standard deviation function (STDEV.S) rather than the population function.

J4 (K-S Test Statistic, D): $= \text{MAX}(G2:G21)$. This final formula extracts the single largest deviation calculated across all observations in Column G, thereby providing the final K-S test statistic (D).

Step 3: Interpretation and Hypothesis Testing

The final and most crucial phase involves interpreting the calculated K-S test statistic (D) by comparing it against a predetermined **critical value**. This comparison allows us to make a formal decision regarding the distributional assumption of the data. The test is set up around the following two formal hypotheses:

H0 (Null Hypothesis): The data follows a **Normal Distribution**.

HA (Alternative Hypothesis): The data does not follow a normal distribution.

From our calculations performed in Excel (specifically Cell J4), the calculated **Maximum D**

statistic is 0.10983. This numerical value quantifies the maximum observed vertical distance between the cumulative probabilities derived from the sample data and the cumulative probabilities that would be expected if the data perfectly adhered to a normal distribution.

To establish statistical significance, this D value must be rigorously assessed against the corresponding **critical value**, which is typically sourced from a standardized Kolmogorov-Smirnov table. The critical value is fundamentally dependent on two factors: the sample size (n) and the chosen significance level (α). Assuming the standard significance level of $\alpha = 0.05$ and our sample size of $n = 20$, the relevant critical value for the K-S test is approximately 0.294.

The established decision rule in hypothesis testing dictates that we must reject the **null hypothesis** (H_0) if the calculated D statistic is greater than the critical value. Conversely, if the calculated D is smaller than or equal to the critical value, we are compelled to fail to reject the null hypothesis, meaning we lack sufficient evidence to conclude non-normality.

Applying this rule to our results, since our calculated D statistic (**0.10983**) is demonstrably less than the critical value (0.294), we formally **fail to reject the null hypothesis** (H_0). This highly favorable conclusion provides statistical confirmation that there is insufficient evidence to suggest the sample distribution deviates significantly from a normal distribution. Consequently, researchers can confidently proceed with parametric statistical tests, knowing that the foundational assumption of normality has been met for this particular dataset.

Conclusion and Next Steps in Statistical Analysis

Successfully executing the K-S test in Excel is a vital step toward ensuring that all foundational distributional assumptions are met prior to engaging in sophisticated statistical modeling. This rigor is paramount for reliable research outcomes. Should the K-S test result in the rejection of the null hypothesis--indicating significant non-normality--analysts would then be required to explore specific non-parametric statistical alternatives or apply appropriate data transformation techniques (such as logarithmic or square-root transformations) in an effort to achieve the required normality.

The following tutorials provide guidance on how to perform other common and essential statistical tests directly within the Microsoft Excel environment, building upon the foundational knowledge gained from mastering the Kolmogorov-Smirnov test: