

How do I perform a statistical comparison of two datasets?

Authored by
stats writer

November 18, 2025

RECOMMENDED CITATION

stats writer (2025). *How do I perform a statistical comparison of two datasets?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=95985>

In the world of data analysis, it is frequently necessary to perform a statistical comparison of two distinct datasets. Whether you are analyzing sales figures from two quarters, examining test scores from two different student groups, or evaluating performance metrics before and after an intervention, understanding the differences in the underlying distributions is paramount. Merely looking at the raw numbers often fails to convey the true story; robust statistical methods are required to quantify differences in central tendency, variation, and spread.

This comprehensive guide details two primary, accessible methods for achieving a robust statistical comparison directly within Microsoft Excel. These techniques move beyond simple visual inspection, providing quantifiable metrics that allow for objective decision-making. We will explore both the detailed analysis offered by the Five number summary and the efficiency of comparing measures of central location and dispersion, ensuring you can choose the right tool for your specific analytical needs.

By the end of this article, you will be equipped to calculate essential statistical measures in Excel, interpret the results accurately, and confidently draw meaningful conclusions about how the distribution of values differs between any two given datasets. Mastery of these methods forms the foundation of effective comparative data analysis, enabling analysts to identify subtle yet significant distributional shifts.

Method 1: Calculating the Five-Number Summary of Each Dataset

The Five number summary (FNS) is a highly effective, non-parametric approach used to summarize the distribution of a set of data. It is the core data required to construct a box plot, offering a comprehensive view of the dataset's shape, central tendency, and variability. This method is particularly valuable because it is less sensitive to outliers compared to methods relying solely on the mean and standard deviation, providing a robust picture of the data's core structure.

The FNS consists of five specific descriptive statistics that define the boundaries and key central points of the data distribution. By calculating these five values for each of your two datasets, you can immediately gain insights into whether the data ranges overlap, where the bulk of the data lies, and whether the distribution is symmetric or skewed. Comparing these five components side-by-side offers a clear, structured framework for comparison.

The components of the five-number summary are defined as follows:

The **minimum value**: The smallest observation in the dataset.

The **first quartile** (Q1 or 25th percentile): The value below which 25% of the data falls.

The **median** (Q2 or 50th percentile): The central value separating the upper half from the lower half of the data. This is a measure of central tendency.

The **third quartile** (Q3 or 75th percentile): The value below which 75% of the data falls.

The **maximum value**: The largest observation in the dataset.

The difference between the first and third quartiles yields the Interquartile Range (IQR), which measures the spread of the middle 50% of the data. This metric is a powerful tool for comparing the variability between datasets without being unduly influenced by extreme outlying observations.

Method 2: Calculating the Average and Standard Deviation

A more traditional and often simpler way to perform a statistical comparison, especially for datasets that are assumed to be normally distributed, is by calculating measures of central tendency and statistical dispersion. This method relies primarily on two core metrics: the mean (average) and the standard deviation.

The **average** (or mean) provides the central location of the dataset. Comparing the averages of two datasets quickly informs us about which dataset generally possesses higher or lower values. However, the average is susceptible to extreme values (outliers), which can distort the perception of the dataset's center. It is crucial to use the mean in conjunction with a measure of spread to paint a complete picture.

The standard deviation (StDev) is the definitive measure of how spread out the values are from the average. A higher standard deviation indicates that the data points are widely dispersed over a large range of values, while a smaller standard deviation suggests that the data points tend to be very close to the mean. By calculating and comparing the standard deviations of both datasets, we can immediately understand which group exhibits greater variability or consistency.

This two-metric approach is computationally efficient and widely understood. It helps us understand roughly where the "center" value is located and how spread out the values are in each dataset. For many preliminary analyses, this method provides sufficient data to make initial comparative assessments, particularly when the sample sizes are similar and the data distributions do not exhibit extreme skewness.

Example: Perform Statistical Comparison of Two Datasets in Excel

To illustrate these two methods in practice, let us examine a hypothetical scenario involving academic performance. Suppose we have collected two separate datasets in Excel, representing the exam scores received by students in two distinct classes, Class 1 and Class 2, on a standardized examination. Our goal is to determine if there are significant differences in performance between the two groups using statistical measures.

The data is structured in two adjacent columns, making it easy for Excel's built-in functions to process the ranges independently. Class 1 scores are listed in Column A, and Class 2 scores are

listed in Column B. The following visualization shows the initial setup of our raw data:

	A	B	C	D	E	F
1	Class 1	Class 2				
2	65	71				
3	66	71				
4	68	72				
5	70	72				
6	71	74				
7	71	75				
8	73	75				
9	76	78				
10	80	80				
11	81	81				
12	81	81				
13	82	83				
14	84	83				
15	88	84				
16	90	84				
17	91	85				
18	92	88				
19	94	88				
20	94	89				
21	96	91				
22						
23						

Note that both datasets contain 20 data points (from row 2 to row 21), ensuring that the comparison is made between samples of equal size. We will now use an adjacent column (Column E) to calculate and organize the statistical summaries for both classes, allowing for clear, direct comparison across the rows.

Calculating the Five-Number Summary in Excel

We will begin by applying the functions necessary to generate the Five-Number Summary for Class 1 (data range A2:A21). Excel provides specific, powerful functions for each component of the FNS, making the calculation straightforward.

We type the following formulas into cells in column E to calculate the five-number summary of the exam scores for Class 1. Note the use of the QUARTILE function, which requires both the array (data range) and a numerical argument (1 for Q1, 2 for Median, 3 for Q3) to specify the desired quartile:

E2 (Minimum): =MIN(A2:A21)

E3 (First Quartile, Q1): =QUARTILE(A2:A21, 1)

E4 (Median): =MEDIAN(A2:A21)

E5 (Third Quartile, Q3): =QUARTILE(A2:A21, 3)

E6 (Maximum): =MAX(A2:A21)

After calculating these five values for Class 1, the power of Excel allows for rapid replication. We can select the range E2:E6 and use the "fill handle" (click and drag) to extend these formulas horizontally to column F, automatically adjusting the range reference from A2:A21 to B2:B21. This step instantly generates the corresponding Five-Number Summary for Class 2.

The resulting table clearly juxtaposes the distributions, highlighting the minimums, maximums, and critical quartiles, which are essential for visual comparisons using box plots or numerical interpretations of spread. Observing the differences in the IQR and range directly from this table already provides valuable insights into relative data variability.

	A	B	C	D	E	F
1	Class 1	Class 2			Class 1	Class 2
2	65	71		Min	65	71
3	66	71		1st Quartile	71	74.75
4	68	72		Median	81	81
5	70	72		3rd Quartile	90.25	84.25
6	71	74		Max	96	91
7	71	75				
8	73	75				
9	76	78				
10	80	80				
11	81	81				
12	81	81				
13	82	83				
14	84	83				
15	88	84				
16	90	84				
17	91	85				
18	92	88				
19	94	88				
20	94	89				
21	96	91				
22						
23						

Calculating Central Tendency and Dispersion in Excel

Next, we apply Method 2 by calculating the arithmetic mean (average) and the standard deviation for both classes. These calculations are typically placed below the FNS summary or in a separate, clearly labeled section for distinction.

We use the following Excel functions for Class 1 (A2:A21). We use the modern standard deviation function for samples (STDEV.S), which is the most common use case for comparative statistics:

E8 (Average/Mean): =AVERAGE (A2 : A21)

E9 (Standard Deviation): =STDEV.S (A2 : A21)

We drag these two formulas from cells E8:E9 to the right into F8:F9. This automatically calculates the mean and standard deviation for Class 2 (B2:B21). These two metrics provide a concise summary of the distribution's center and overall spread.

The final summary table, combining both the five-number summary and the mean/standard deviation, is essential for a holistic statistical review. It allows the analyst to cross-reference the median (a robust measure of center) with the mean (a classical measure of center) and compare the Interquartile Range (IQR, a robust measure of spread) with the standard deviation (a classical measure of spread).

E8 \times \checkmark f_x =AVERAGE(A2:A21)							
	A	B	C	D	E	F	G
1	Class 1	Class 2			Class 1	Class 2	
2	65	71		Min	65	71	
3	66	71		1st Quartile	71	74.75	
4	68	72		Median	81	81	
5	70	72		3rd Quartile	90.25	84.25	
6	71	74		Max	96	91	
7	71	75					
8	73	75		Mean	80.65	80.25	
9	76	78		Std. Deviation	10.21493	6.430806	
10	80	80					
11	81	81					
12	81	81					
13	82	83					
14	84	83					
15	88	84					
16	90	84					
17	91	85					
18	92	88					
19	94	88					
20	94	89					
21	96	91					
22							

Conclusion 1: Analyzing the Central Tendency (Center)

A crucial first step in any statistical comparison is assessing the central tendency, or the typical value, of each dataset. In our example, we analyze two metrics of center: the median and the mean. These two metrics often diverge if the data is skewed, but their proximity indicates a relatively symmetrical distribution.

Upon reviewing the calculated values, we observe that both datasets share an identical **median** exam score of 81. This metric tells us that half the students in Class 1 scored 81 or above, and half the students in Class 2 also scored 81 or above. The median, being resistant to extreme scores, strongly suggests that the typical student performance in both classes is remarkably similar.

Furthermore, the **mean** values are only slightly different. Class 1 has an average exam score of 80.65, while Class 2 has an average exam score of 80.25. The marginal difference of 0.4 points is statistically insignificant in this context. The close alignment of both the mean and the median reinforces the conclusion that the central or "typical" exam score between the two classes is almost identical. If the mean had been significantly lower or higher than the median in one class, it would

have signaled skewness or the presence of significant outliers, demanding further investigation.

Therefore, we conclude that based on measures of central location, there is no meaningful difference in the overall performance level between Class 1 and Class 2. The variation, however, presents a very different story, which we address in the next section.

Conclusion 2: Analyzing Variability and Spread

While the central tendency was similar, the measures of dispersion--the metrics that quantify how spread out the scores are--reveal a stark contrast between the two classes. Analyzing the variability is often more informative than analyzing the center, as it speaks directly to the consistency and distribution of performance. Our analysis focuses on three key metrics: the range, the interquartile range (IQR), and the standard deviation.

First, consider the **Range** (Maximum minus Minimum). The range provides the simplest measure of overall spread, encompassing all data points:

Range of Class 1: $96 - 65 = 31$

Range of Class 2: $91 - 71 = 20$

The range for Class 1 is significantly larger (31 compared to 20), indicating that the scores in Class 1 span a wider spectrum, containing both lower and higher extremes than Class 2. This suggests less consistency in performance within Class 1.

Next, we examine the Interquartile Range (IQR = $Q3 - Q1$), which measures the spread of the middle 50% of the data and is less sensitive to the extreme outliers at the minimum and maximum ends:

Interquartile Range of Class 1: $90.25 - 71 = 19.25$

Interquartile Range of Class 2: $84.25 - 74.75 = 9.5$

The IQR for Class 1 (19.25) is more than double that of Class 2 (9.5). This is a powerful indication that even excluding the most extreme 25% on either side, the core group of students in Class 1 exhibits much greater heterogeneity in their scores compared to the core group in Class 2, which is highly clustered around the median.

Finally, the **Standard Deviation** (StDev) quantifies the average amount of variation or deviation from the mean, providing the most commonly cited measure of dispersion:

Standard Deviation of Class 1: **10.21**

Standard Deviation of Class 2: **6.43**

The standard deviation for Class 1 is substantially higher (10.21 vs. 6.43). This confirms the findings from the range and IQR: the scores in Class 1 are much more widely dispersed around their average than the scores in Class 2. Students in Class 2 demonstrated much more consistent performance relative to their class average.

Summary of Findings and Analytical Implications

The statistical comparison clearly demonstrates that while both classes achieved roughly the same typical performance level (central tendency), Class 1 exhibits significantly higher variability in its scores. This suggests different instructional outcomes or perhaps different demographic compositions within the two classes. Class 2 shows consistent, tightly clustered performance, whereas Class 1 contains a wider mix of high and low achievers.

When selecting which statistical measure to rely on, it is important to consider the underlying distribution. If the data shows significant skewness or extreme outliers, the Five number summary (especially the median and IQR) provides a more robust and representative summary of the data structure. If the distribution is roughly symmetrical and free of major outliers, the mean and standard deviation are excellent, efficient metrics for comparison.

For analysts seeking to move beyond descriptive statistics into inferential analysis, these initial findings lay the groundwork for formal hypothesis testing. For instance, since the means are very close, one might perform a two-sample T-test to determine if the small difference in the means (80.65 vs. 80.25) is statistically significant, taking into account the large difference in variances (as indicated by the standard deviation). However, the descriptive statistics alone--especially the variance metrics--already provide profound operational insights into the comparative performance of the two groups.

Mastering these two methods in Excel allows for quick, actionable, and statistically sound comparisons of any two datasets, moving you from raw data to informed conclusions with confidence.