

Understanding Percentiles in Excel: A Comparison of PERCENTILE.EXC and PERCENTILE.INC

Authored by
Mohammed loot

November 3, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Understanding Percentiles in Excel: A Comparison of PERCENTILE.EXC and PERCENTILE.INC*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=9477>

The n th [percentile](#) is a fundamental metric within descriptive statistics, employed extensively across diverse fields ranging from financial risk assessment to standardized educational testing. Conceptually, the n th **percentile** of a sorted [dataset](#) defines the value below which n percent of the recorded observations are found. This calculation relies on the prerequisite that all data points must be meticulously ordered from the smallest magnitude to the largest, providing a clear map of the distribution's shape.

To provide a tangible example, if a job applicant scores in the 90th percentile on an aptitude test, this means the achieved score is equal to or greater than 90% of all other scores recorded within the reference group. Percentiles offer a highly granular measure of relative standing, distinguishing them from broader measures like quartiles or deciles, which divide the data into specific, fixed segments such as quarters or tenths. This granularity makes percentiles a powerful tool for comparative analysis.

While the basic statistical definition of a percentile is consistent, the precise computational methodology used to locate that percentile can vary slightly, particularly when the calculated rank position falls between two existing data points. This methodological ambiguity--concerning how boundary values are treated and how [interpolation](#) is performed--is precisely what necessitates the distinction between Excel's specialized functions, **PERCENTILE.EXC** and **PERCENTILE.INC**, making their differences crucial for accurate data analysis.

Understanding Excel's Three Percentile Functions

Microsoft Excel offers three distinct functions designed to calculate the k th percentile of an array or range of numerical data. Although their statistical objective is identical, the underlying mathematical boundary conditions they employ lead to measurable differences in results, especially when analysts are working with datasets containing a small number of observations. Ensuring alignment with external statistical packages requires a clear understanding of these function variations.

The three available functions, classified by their approach to boundary handling, are:

PERCENTILE.EXC: This function calculates the k th percentile using an **exclusive** method. This means it strictly excludes the extreme boundary values of 0 and 1 (or 0% and 100%). Therefore, the permissible range for the input k must be the open interval $(0, 1)$ --values strictly between 0 and 100%. This methodological standard is frequently referred to as the R-6 method in established statistical literature. The use of this function guarantees that the resulting percentile value will never equal the minimum or maximum observation in the dataset.

PERCENTILE.INC: This function calculates the k th percentile using an **inclusive** method. This is achieved by including the boundary values of 0 and 1 (0% and 100%) within the calculation range. The permissible range for k is therefore the closed interval $[0, 1]$. If $k=0$ is specified, the function returns

the minimum value of the array; if $k=1$, it returns the maximum value. This inclusive approach aligns closely with the long-standing industry definition, often labeled the R-7 method.

PERCENTILE: This function serves primarily as a legacy function, retained solely for backward compatibility with documents created in much older versions of Excel. It is essential for analysts to note that the **PERCENTILE** function operates using the exact same calculation methodology as the modern **PERCENTILE.INC** function. Microsoft strongly recommends that analysts utilize the explicit **.INC** or **.EXC** suffixes in new analyses to clearly communicate the intended method of calculation.

The choice of function dictates how the tool handles the [interpolation](#) of values when the desired percentile rank does not perfectly correspond to an existing observation within the input array, thereby influencing the final statistical result.

Technical Differences in Rank Calculation and Boundary Logic

The core divergence between the inclusive (**.INC**) and exclusive (**.EXC**) methods fundamentally rests on their approach to calculating the specific rank position (i) within a sorted dataset of N values. Since i is often fractional, both methods rely on linear interpolation to estimate the final percentile value, but the formulas used to derive the initial rank i are mathematically distinct.

For the inclusive method (**PERCENTILE.INC**), which treats the data as a complete statistical population, the rank position (i) is determined using the formula: $i = k * (N - 1) + 1$. Because the calculation incorporates N minus 1, the total possible index range spans from 1 up to N . This structural design ensures that the calculated percentile value can precisely match the minimum value (when $k=0$) or the maximum value (when $k=1$) observed within the array, fulfilling the requirement of the R-7 methodology.

In contrast, the exclusive method (**PERCENTILE.EXC**), often used when treating the data as a sample to estimate population parameters, calculates the rank position using the formula: $i = k * (N + 1)$. Since the input k must be strictly between 0 and 1, the resulting position i will always fall strictly between the first and last observation's index. This inherent design ensures that the **PERCENTILE.EXC** function will never return the absolute minimum or maximum values of the dataset, effectively avoiding the boundary observations.

When the calculated rank i is not a whole number, Excel employs linear interpolation. The rank i is partitioned into an integer part (J , which identifies the lower index position) and a fractional part (G , which serves as the interpolation weight). The resulting percentile value (P) is then determined by the formula: $P = V(J) + G * (V(J+1) - V(J))$, where V is the sorted data array. Because the initial rank calculation formulas for **.INC** and **.EXC** inherently produce slightly different values for i --and thus different fractional parts G --they inevitably yield distinct final interpolated percentile results. This mathematical shift in the rank position is the technical cause of the differences observed

between the two functions.

Practical Demonstration with an Excel Dataset

To fully appreciate how these differing boundary assumptions affect real-world outcomes, we will examine a controlled, small [dataset](#). Suppose we have the following sample data points (N=10 observations) located in cells A1 through A10 in an Excel spreadsheet:

	A	B	C	D	E	F	G	H
1	Data							
2	2							
3	4							
4	5							
5	6							
6	6							
7	7							
8	9							
9	12							
10	14							
11	15							
12	18							
13	19							
14	22							
15	24							
16	26							
17	28							
18								
19								
20								
21								
22								
23								
24								

Our objective is to calculate the 20th [percentile](#) ($k=0.2$) for this numerical array.

The following screenshot demonstrates the application of the three functions and the resulting calculations when seeking the 20th percentile:

	A	B	C	D	E	F
1	Data			20th percentile	Formula	
2	2		PERCENTILE	6	=PERCENTILE(A2:A17, 0.2)	
3	4		PERCENTILE.INC	6	=PERCENTILE.INC(A2:A17, 0.2)	
4	5		PERCENTILE.EXC	5.4	=PERCENTILE.EXC(A2:A17, 0.2)	
5	6					
6	6					
7	7					
8	9					
9	12					
10	14					
11	15					
12	18					
13	19					
14	22					
15	24					
16	26					
17	28					
18						
19						
20						
21						
22						
23						
24						
25						
26						

Using the inclusive methods (**PERCENTILE.INC** or **PERCENTILE**), the resulting 20th percentile is calculated as 6. This figure is derived from the R-7 calculation where the position $i = 0.2 * (10 - 1) + 1 = 2.8$. Since the rank is 2.8, interpolation occurs between the 2nd and 3rd sorted data points, yielding 6 as the final value.

Conversely, using the exclusive method (**PERCENTILE.EXC**), the resulting 20th percentile is calculated as 5.4. This outcome is derived from the R-6 calculation where the position $i = 0.2 * (10 + 1) = 2.2$. While this also requires interpolation between the 2nd and 3rd sorted values, the position (2.2) is shifted closer to the 2nd point than the 2.8 position, demonstrating how the boundary assumptions fundamentally shift the final result.

Industry Standards and Cross-Platform Consistency

In professional statistical practice, where data integrity often requires replication and sharing across different software environments, methodological consistency is of utmost importance. For this reason, **PERCENTILE.INC** is typically the preferred function in statistical analysis and is largely considered the de facto standard definition of a percentile due to its robust alignment with

established practices.

This preference is heavily influenced by the implementations used in popular statistical programming environments. Specifically, the calculation methods utilized by the standard libraries in both the [R programming language](#) and the data analysis packages (such as NumPy and Pandas) within the [Python programming language](#) default to algorithms that closely match the inclusive definition (R-7). Consequently, employing **PERCENTILE.INC** in Excel ensures that percentile calculations performed within your spreadsheet will be consistent with results generated by advanced statistical software, minimizing discrepancies during cross-platform reporting and validation.

While the numerical difference between the **.INC** and **.EXC** results may be statistically negligible for very large datasets ($N > 1000$), the divergence can be highly significant when working with smaller samples (e.g., $N < 50$). For general business intelligence, financial modeling, or academic research requiring strict adherence to the standard statistical definition, the inclusive function is the consistently recommended choice.

Summary of Method Selection and Best Practices

The selection of the correct percentile function in Excel must be guided by the specific statistical context of the analysis and the required treatment of boundary conditions. For nearly all practical business and scientific applications where the dataset is viewed as a population sample or where consistency with widely accepted statistical software is required, **PERCENTILE.INC** is the most suitable and statistically robust choice.

Key takeaways for choosing your function:

Select **PERCENTILE.INC** (or its legacy alias **PERCENTILE**) when the analysis requires the resulting percentile to potentially equal the minimum (0%) or maximum (100%) value observed in the dataset. This adheres to the standard R-7 method.

Use **PERCENTILE.EXC** only when the statistical intent is to define the percentile boundaries as strictly internal to the observed data range, meaning the calculation must never return the minimum or maximum values. This aligns with the R-6 method.

It is critical to always document which function you have used within your data analysis models and reports to ensure clarity, auditability, and reproducibility of your findings.

A strong understanding of these technical nuances is essential for any professional data analyst, ensuring that statistical reports are accurate, mathematically defensible, and consistent with broader industry standards.