

How to use Linear Interpolation in Excel?

Authored by
stats writer

December 16, 2025

RECOMMENDED CITATION

stats writer (2025). *How to use Linear Interpolation in Excel?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=107624>

Linear interpolation in Excel is an immensely powerful technique used by analysts and engineers to accurately estimate missing data points within a known range of observations. This mathematical technique assumes a linear relationship between adjacent known data points, allowing for the calculation of an unknown value that resides directly between them. The primary utility of this method is filling gaps in time-series data, environmental readings, or financial models where continuous observation was impossible or impractical.

To successfully implement linear interpolation in a spreadsheet environment, you must first precisely define the two bounding data pairs--the points above and below the unknown value--that will anchor your calculation. Once identified, the process relies on applying the fundamental principles derived from the slope-intercept form of a straight line equation. This calculation determines the gradient of the line segment connecting the known points and projects that trend to the target unknown x-value.

The final step involves synthesizing this mathematical logic into a cohesive and efficient Excel formula. While a simple interpolation might be achievable manually, advanced applications often require dynamic formulas--incorporating functions like MATCH and OFFSET--to locate the relevant known points automatically. This tutorial provides a comprehensive guide to implementing this dynamic and robust interpolation method.

The Mathematical Foundation of Linear Interpolation

Interpolation is formally defined as the general procedure of constructing new data points within the range of a discrete set of known data points. Specifically, linear interpolation is the simplest form, using a straight line to connect the nearest known neighbors on either side of the point to be estimated. This method is effective when the underlying data trend is relatively smooth or slowly changing.

When we are given two defined coordinate pairs, (x_1, y_1) and (x_2, y_2) , our goal is to estimate the corresponding y-value for any given intermediate x-value. This estimation relies entirely on the principle of similar triangles formed by the line segment and the horizontal and vertical axes, which translates mathematically into the linear equation. The accuracy of the interpolated value depends on how closely the true relationship between the variables adheres to a straight line segment between those two points.

The core formula that governs linear interpolation is a rearrangement of the slope-intercept form, often expressed to solve for the unknown y-value (y) corresponding to the new x-value (x). The formula effectively calculates the slope, multiplies it by the distance between the starting x-point and the new x-point, and adds the result to the starting y-value:

$$y = y_1 + (x - x_1) \cdot \frac{y_2 - y_1}{x_2 - x_1}$$

This tutorial explains how to integrate this powerful mathematical concept into a dynamic and automated formula within Excel, allowing users to find any unknown y-value based on an x-value without manually identifying the bounding points every time.

Preparing the Dataset for Interpolation

To demonstrate the practical application of linear interpolation, let us consider a sample dataset. This data might represent measured sensor outputs, temperature readings over time, or production yield across different input levels. For dynamic interpolation to function correctly in Excel, it is critical that the independent variable (X-values) be sorted in ascending order. Our example assumes this foundational requirement is met.

Suppose we have the following sample dataset structured in two columns in Excel, where the first column contains the known X-values and the second contains the corresponding Y-values:

	A	B	C	D	E	F
1	x	y				
2	2	4				
3	4	7				
4	6	11				
5	8	16				
6	10	22				
7	12	29				
8	14	38				
9	16	49				
10	18	63				
11	20	80				
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						

This simple table forms the basis of our interpolation effort. The X-values range from 10 to 20, and the Y-values show a clear, though not perfectly linear, increasing trend. Having the data clearly labeled and organized is the essential first step before attempting any complex calculation.

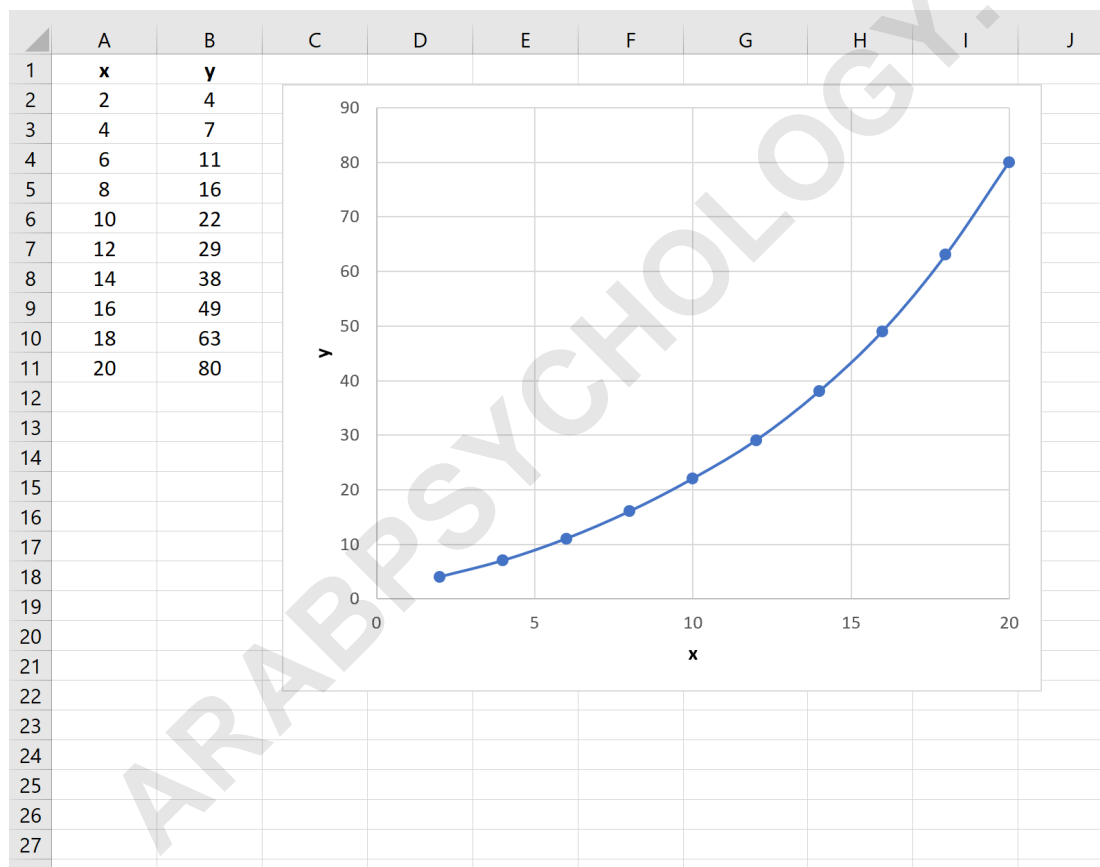
Visualizing the data is often the best way to understand the trend and confirm that linear

interpolation is an appropriate methodology. If the plotted points exhibited extreme curvature or oscillation, a different technique, such as polynomial or spline interpolation, might be necessary. However, for most basic estimations, the linear approach suffices.

Visualizing the Data Trend

Creating a quick scatter plot or line chart of the existing data allows us to instantly visualize the relationship between our independent and dependent variables. This graphical representation confirms the increasing trend and highlights the segments where the interpolation will take place.

If we create a quick plot of the data points from our table, here's what the visual representation of our known data points looks like:



As shown in the plot, the relationship between X and Y is largely positive. The line segments connecting the known points are nearly straight, confirming that treating the relationship as linear within short intervals--the core principle of linear interpolation--will yield highly accurate estimates.

The goal now is to introduce a new X-value that falls within the range of this existing data and calculate its corresponding Y-value based on the straight line drawn between its two nearest neighbors. This process of visual confirmation ensures that the mathematical model we are about

to apply is statistically sensible for the given data structure.

Defining the Missing Data Point Challenge

Our specific objective in this example is to find the Y-value associated with a new, unobserved X-value of **13**. Looking at our initial dataset, we can clearly see that we have defined, *measured* Y-values for X-values of 12 (Y=32) and 14 (Y=35). However, the value for X=13 is missing.

In the context of the linear interpolation formula, our known bounding points are $(x_1, y_1) = (12, 32)$ and $(x_2, y_2) = (14, 35)$. Our target X-value is 13. Given the simplicity of this particular case, we can manually deduce that the resulting Y-value should be halfway between 32 and 35, which is 33.5.

However, for large datasets where the target X-value is not a simple midpoint, or where we need to run multiple interpolations quickly, manually identifying x_1 , y_1 , x_2 , and y_2 is inefficient and prone to error. This is why we must develop a single, dynamic Excel formula that automatically locates the correct bounding points for any X-value entered, provided it falls within the overall range.

Developing the Dynamic Excel Interpolation Formula

To perform linear interpolation dynamically in Excel, we must combine several functions. The core calculation utilizes the FORECAST function (or FORECAST.LINEAR in newer versions), which is designed to calculate a future value along a linear trend. However, FORECAST needs only the two immediate bounding points, not the entire range.

We use the MATCH function to identify the position of the known X-value that is immediately less than or equal to our new target X-value. By setting the match type to 1 (less than or equal to), we find the starting point (x_1) of our interpolation segment.

Next, the OFFSET function is used twice--once for the X-values and once for the Y-values. OFFSET takes the starting position found by MATCH and creates a two-cell range (the starting point and the point immediately following it), defining both the (x_1, y_1) and (x_2, y_2) pairs that the FORECAST function requires.

The final concatenated formula, which encapsulates all this logic, is presented below. Note that *NewX*, *KnownY*, and *KnownX* refer to the cell containing the target X-value (13), the range of known Y-values, and the range of known X-values, respectively:

```
=FORECAST(NewX,OFFSET(KnownY,MATCH(NewX,KnownX,1)-1,0,2),  
OFFSET(KnownX,MATCH(NewX,KnownX,1)-1,0,2))
```

Step-by-Step Implementation in Excel

Implementing this complex formula requires precision in cell referencing. Assuming the X-values (KnownX) are in cells A2:A7 and the Y-values (KnownY) are in cells B2:B7, and the target X-value (NewX=13) is entered into cell D2, we can substitute the range names with actual cell references.

First, the `MATCH(D2, A2:A7, 1)` part identifies the largest value in A2:A7 that is less than or equal to 13. This correctly returns the position of X=12. We subtract 1 from this result because `OFFSET` calculates from the starting cell (A2 or B2).

Next, the two `OFFSET` functions dynamically define the ranges for the known Ys and known Xs. For Ys, the formula creates the range {32; 35}, and for Xs, it creates the range {12; 14}. These two short ranges are passed to the `FORECAST` function, which then calculates the interpolated value for the target X=13 based solely on those two points.

Here's how the application looks within the spreadsheet environment, where the formula is entered into cell E2 to estimate the y-values associated with an x-value of 13:

	A	B	C	D	E	F	G	H	I	J	K	L
1	x	y										
2	2	4										
3	4	7										
4	6	11										
5	8	16										
6	10	22										
7	12	29										
8	14	38										
9	16	49										
10	18	63										
11	20	80										
12												
13	=FORECAST(13,OFFSET(B2:B11,MATCH(13,A2:A11,1)-1,0,2), OFFSET(A2:A11,MATCH(13,A2:A11,1)-1,0,2))											
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												

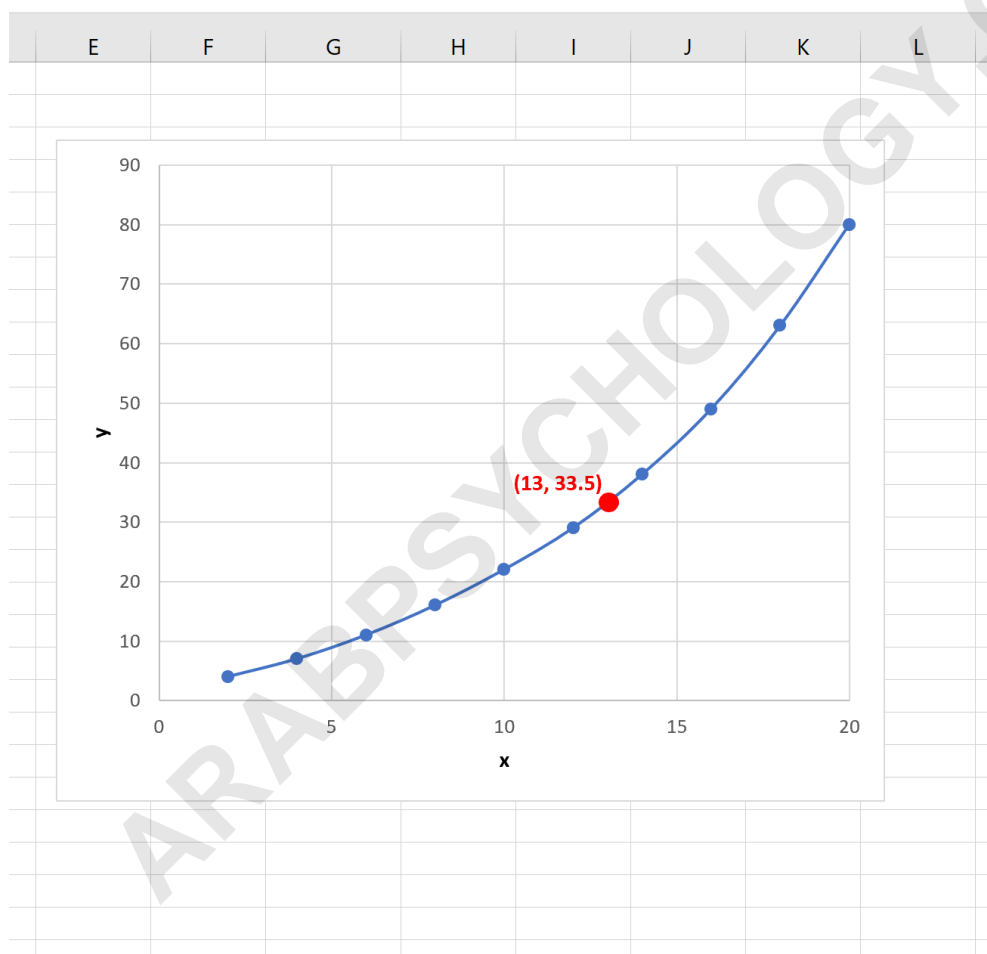
Analyzing the Interpolation Results and Visual Validation

Upon execution of the dynamic formula in Excel, the estimated y-value returned is precisely **33.5**. This result mathematically confirms our initial expectation, as 13 is the exact midpoint between 12

and 14, and 33.5 is the exact midpoint between 32 and 35.

The power of this method is evident when considering more complex scenarios where the target X-value is not a simple midpoint (e.g., finding the Y for $X=12.3$). In such cases, the formula automatically applies the correct weighting based on the distance from the known points, providing an accurate estimate based on the local slope calculated by FORECAST.

To visually confirm the accuracy of our interpolation, we can add the newly calculated point (13, 33.5) to our existing plot. If the interpolation was successful and the linear assumption holds true for this segment, the new point should fall perfectly onto the line segment connecting (12, 32) and (14, 35).



As clearly demonstrated in the updated chart, the interpolated point sits perfectly along the established trend line between the surrounding measured points. This visual verification provides high confidence in the accuracy of the linear interpolation result derived from the complex Excel formula.

Important Limitations and Considerations

While this dynamic formula is robust, it is crucial to understand its limitations. The primary constraint of linear interpolation is that the new x-value must strictly fall **within the range** of the existing x-values. This process is called interpolation, meaning "between the points."

If you attempt to find a Y-value for an X-value that falls outside the boundaries of your known dataset (e.g., X=5 or X=25), you would be performing **extrapolation**. While the FORECAST function is capable of extrapolation, the resulting values are generally far less reliable than interpolated values, as there is no bounding point to constrain the calculation. Extrapolation assumes the linear trend continues indefinitely, which is often mathematically unsound in real-world data.

Furthermore, this method assumes that the known X-values are sorted ascendingly. If the data is not sorted, the MATCH function with the match type 1 will fail to correctly identify the immediate bounding points, leading to erroneous results. Always ensure your foundational data is properly ordered before applying this dynamic interpolation technique in Excel.

You can find more advanced Excel tutorials covering data manipulation and statistical functions in our resource library.