

# How can I create a confusion matrix in Excel?

Authored by  
**stats writer**

May 6, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I create a confusion matrix in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143441>

A confusion matrix is a visual representation of the performance of a classification model. It shows the number of correct and incorrect predictions made by the model for each class. To create a confusion matrix in Excel, start by organizing your data into four categories: true positives, false positives, true negatives, and false negatives. Then, use the COUNTIF function to count the number of observations in each category. Next, create a table with the four categories as rows and columns. Finally, use conditional formatting to color code the cells based on the number of observations, making it easier to interpret the results. By following these steps, you can effectively create a confusion matrix in Excel to evaluate the accuracy of your classification model.

## Create a Confusion Matrix in Excel

is a type of regression we can use when the response variable is binary.

One common way to evaluate the quality of a logistic regression model is to create a confusion matrix, which is a 2x2 table that shows the predicted values from the model vs. the actual values from the test dataset.

		Predicted	
		0	1
Actual	0	30	12
	1	8	56

The following step-by-step example shows how to create a confusion matrix in Excel.

### Step 1: Enter the Data

**First, let's enter a column of actual values for a response variable along with the predicted values by a logistic regression model:**

	A	B	C	D	E	F	G
1	<b>Actual</b>	<b>Predicted</b>					
2	0	0					
3	0	0					
4	0	1					
5	0	0					
6	0	0					
7	0	1					
8	0	1					
9	0	0					
10	0	0					
11	0	1					
12	1	0					
13	1	0					
14	1	1					
15	1	1					
16	1	1					
17	1	1					
18	1	1					
19	1	1					
20	1	1					
21	1	1					
22							
23							
24							

### Step 2: Create the Confusion Matrix

**Next, we'll use the COUNTIFS() formula to count the number of values that are "0" in the Actual column and also "0" in the Predicted column:**

F3								
=COUNTIFS(A2:A21, "=0", B2:B21, "=0")								
	A	B	C	D	E	F	G	H
1	Actual	Predicted				Predicted		
2	0	0				0	1	
3	0	0		Actual	0	6		
4	0	1			1			
5	0	0						
6	0	0						
7	0	1						
8	0	1						
9	0	0						
10	0	0						
11	0	1						
12	1	0						
13	1	0						
14	1	1						
15	1	1						
16	1	1						
17	1	1						
18	1	1						
19	1	1						
20	1	1						
21	1	1						
22								

We'll use a similar formula to fill in every other cell in the confusion matrix:

	A	B	C	D	E	F	G	H
1	<b>Actual</b>	<b>Predicted</b>				<b>Predicted</b>		
2	0	0				0	1	
3	0	0		<b>Actual</b>		0	6	4
4	0	1				1	2	8
5	0	0						
6	0	0						
7	0	1						
8	0	1						
9	0	0						
10	0	0						
11	0	1						
12	1	0						
13	1	0						
14	1	1						
15	1	1						
16	1	1						
17	1	1						
18	1	1						
19	1	1						
20	1	1						
21	1	1						
22								
23								

### Step 3: Calculate Accuracy, Precision and Recall

Once we've created the confusion matrix, we can calculate the following metrics:

**Accuracy:** Percentage of correct predictions  
**Precision:** Correct positive predictions relative to total positive predictions  
**Recall:** Correct positive predictions relative to total actual positives

The following formulas show how to calculate each of

## these metrics in Excel:

	A	B	C	D	E	F	G	H
1	<b>Actual</b>	<b>Predicted</b>				<b>Predicted</b>		
2	0	0				0	1	
3	0	0		<b>Actual</b>	0	6	4	
4	0	1			1	2	8	
5	0	0						
6	0	0						
7	0	1		<b>Metric</b>	<b>Value</b>	<b>Formula</b>		
8	0	1		Accuracy	0.7	=SUM(F3, G4)/SUM(F3:G4)		
9	0	0		Precision	0.667	=G4/SUM(G3:G4)		
10	0	0		Recall	0.8	=G4/SUM(F4:G4)		
11	0	1						
12	1	0						
13	1	0						
14	1	1						
15	1	1						
16	1	1						
17	1	1						
18	1	1						
19	1	1						
20	1	1						
21	1	1						
22								

In this example, our model has an accuracy of 0.7 which tells us that it correctly classified 70% of observations.

If we'd like, we can compare this accuracy to that of other logistic regression models to determine which model is best at classifying observations into categories of 0 or 1.