

# How can I create a ROC curve in Excel?

Authored by  
**stats writer**

May 5, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I create a ROC curve in Excel?*. PSYCHOLOGICAL SCALES.  
Retrieved from <https://scales.arabpsychology.com/?p=142902>

A ROC (Receiver Operating Characteristic) curve in Excel is a graphical representation of the performance of a binary classifier model. It helps to evaluate the accuracy and effectiveness of a model in predicting the correct class of a binary outcome.

To create a ROC curve in Excel, first, you need to have a dataset with the actual and predicted values of the binary outcome. Then, follow these steps:

1. Sort the data in descending order based on the predicted values.
2. Calculate the True Positive Rate (TPR) and False Positive Rate (FPR) for each predicted value. TPR is the percentage of correctly predicted positive outcomes, while FPR is the percentage of incorrectly predicted negative outcomes.
3. Plot the TPR against the FPR to create the ROC curve.
4. Add a diagonal line from (0,0) to (1,1) to represent the random chance of predicting correctly.
5. The area under the ROC curve (AUC) represents the overall performance of the model, with a higher AUC indicating a better-performing model.

In conclusion, a ROC curve in Excel can be easily created by following these steps, providing a visual representation of the accuracy and effectiveness of a binary classifier model.

## Create a ROC Curve in Excel (Step-by-Step)

**Logistic Regression is a statistical method that we use to fit a regression model when the response variable is binary. To assess how well a logistic regression model fits a dataset, we can look at the following two metrics:**

**Sensitivity: The probability that the model predicts a positive outcome for an observation when indeed the outcome is positive. This is also called the "true positive rate."****Specificity: The probability that the model**

**predicts a negative outcome for an observation when indeed the outcome is negative. This is also called the "true negative rate."**

**One way to visualize these two metrics is by creating a ROC curve, which stands for "receiver operating characteristic" curve. This is a plot that displays the sensitivity and specificity of a logistic regression model.**

**The following step-by-step example shows how to create and interpret a ROC curve in Excel.**

**Step 1: Enter the Data**

**First, let's enter some raw data:**

	A	B	C	D	E	F	G
1		Observed					
2	<b>Data Value</b>	<b>Pass</b>	<b>Fail</b>				
3							
4	< 1	20	4				
5	1 - 1.9	41	5				
6	2 - 2.9	67	7				
7	3 - 3.9	88	12				
8	4 - 4.9	102	14				
9	5 - 5.9	104	18				
10	6 - 6.9	85	34				
11	7 - 7.9	45	33				
12	8 - 8.9	30	35				
13	9 - 9.9	4	22				
14	> 10	2	19				
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							

### Step 2: Calculate the Cumulative Data

Next, let's use the following formula to calculate the cumulative values for the Pass and Fail categories:

**Cumulative Pass values: =SUM(\$B\$3:B3)**

**Cumulative Fail values: =SUM(\$C\$3:C3)**

We'll then copy and paste these formulas down to every cell in column D and column E:

	A	B	C	D	E	F	G
1		Observed		Cumulative			
2	<b>Data Value</b>	<b>Pass</b>	<b>Fail</b>	<b>Pass</b>	<b>Fail</b>		
3				0	0		
4	< 1	20	4	20	4		
5	1 - 1.9	41	5	61	9		
6	2 - 2.9	67	7	128	16		
7	3 - 3.9	88	12	216	28		
8	4 - 4.9	102	14	318	42		
9	5 - 5.9	104	18	422	60		
10	6 - 6.9	85	34	507	94		
11	7 - 7.9	45	33	552	127		
12	8 - 8.9	30	35	582	162		
13	9 - 9.9	4	22	586	184		
14	> 10	2	19	588	203		
15							
16							
17							
18							
19							
20							
21							
22							
23							

### Step 3: Calculate False Positive Rate & True Positive Rate

Next, we'll calculate the false positive rate (FPR), true positive rate (TPR), and the area under the curve (AUC) using the following formulas:

$$\text{FPR: } =1-D3/\$D\$14 \quad \text{TPR: } =1-E3/\$E\$14 \quad \text{AUC: } =(F3-F4)*G3$$

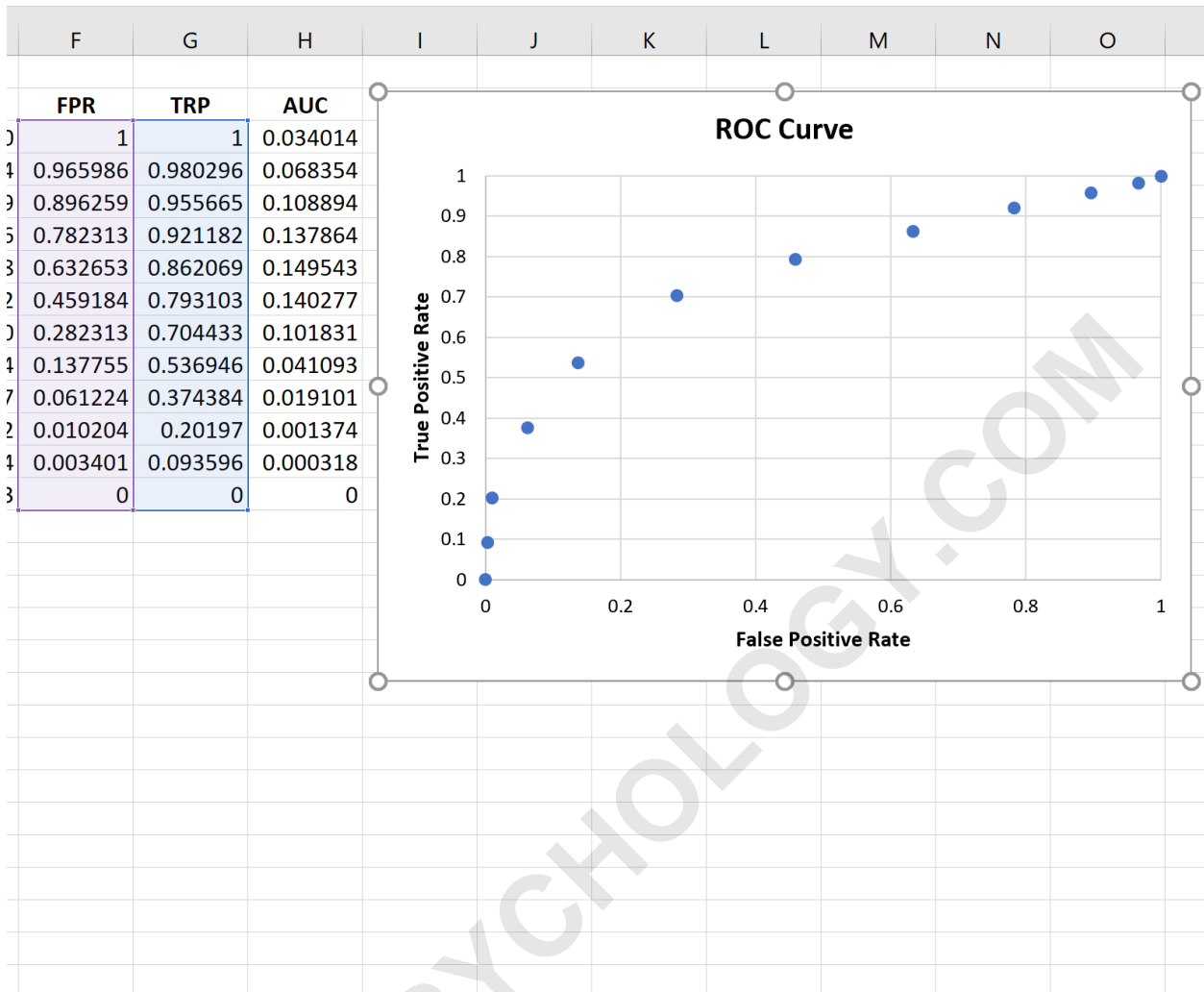
We'll then copy and paste these formulas down to every cell in columns F, G, and H:

	A	B	C	D	E	F	G	H	I
1		Observed		Cumulative					
2	<b>Data Value</b>	<b>Pass</b>	<b>Fail</b>	<b>Pass</b>	<b>Fail</b>	<b>FPR</b>	<b>TRP</b>	<b>AUC</b>	
3				0	0	1	1	0.034014	
4	< 1	20	4	20	4	0.965986	0.980296	0.068354	
5	1 - 1.9	41	5	61	9	0.896259	0.955665	0.108894	
6	2 - 2.9	67	7	128	16	0.782313	0.921182	0.137864	
7	3 - 3.9	88	12	216	28	0.632653	0.862069	0.149543	
8	4 - 4.9	102	14	318	42	0.459184	0.793103	0.140277	
9	5 - 5.9	104	18	422	60	0.282313	0.704433	0.101831	
10	6 - 6.9	85	34	507	94	0.137755	0.536946	0.041093	
11	7 - 7.9	45	33	552	127	0.061224	0.374384	0.019101	
12	8 - 8.9	30	35	582	162	0.010204	0.20197	0.001374	
13	9 - 9.9	4	22	586	184	0.003401	0.093596	0.000318	
14	> 10	2	19	588	203	0	0	0	
15									
16									
17									
18									
19									
20									
21									
22									
23									

#### Step 4: Create the ROC Curve

To create the ROC curve, we'll highlight every value in the range F3:G14.

Then we'll click the Insert tab along the top ribbon and then click Insert Scatter(X, Y) to create the following plot:



**Step 5: Calculate the AUC**

The more that the curve hugs the top left corner of the plot, the better the model does at classifying the data into categories.

As we can see from the plot above, this logistic regression model does a pretty good job of classifying the data into categories.

To quantify this, we can calculate the AUC (area under the curve) which tells us how much of the plot is located under the curve.

The closer AUC is to 1, the better the model. A model with an AUC equal to 0.5 is no better than a model that makes random classifications.

To calculate the AUC of the curve, we can simply take the sum of all of the values in column H:

H15      fx      =SUM(H3:H14)								
	A	B	C	D	E	F	G	H
1		Observed		Cumulative				
2	Data Value	Pass	Fail	Pass	Fail	FPR	TRP	AUC
3				0	0	1	1	0.034014
4	< 1	20	4	20	4	0.965986	0.980296	0.068354
5	1 - 1.9	41	5	61	9	0.896259	0.955665	0.108894
6	2 - 2.9	67	7	128	16	0.782313	0.921182	0.137864
7	3 - 3.9	88	12	216	28	0.632653	0.862069	0.149543
8	4 - 4.9	102	14	318	42	0.459184	0.793103	0.140277
9	5 - 5.9	104	18	422	60	0.282313	0.704433	0.101831
10	6 - 6.9	85	34	507	94	0.137755	0.536946	0.041093
11	7 - 7.9	45	33	552	127	0.061224	0.374384	0.019101
12	8 - 8.9	30	35	582	162	0.010204	0.20197	0.001374
13	9 - 9.9	4	22	586	184	0.003401	0.093596	0.000318
14	> 10	2	19	588	203	0	0	0
15								<b>0.802662</b>
16								
17								
18								
19								
20								
21								
22								

**The AUC turns out to be 0.802662. This value is fairly high, which indicates that the model does a good job of classifying the data into 'Pass' and 'Fail' categories.**

**The following tutorials explain how to create other common plots in Excel:**

ARABPSYCHOLOGY.COM