

# Understanding Confusion Matrices for Logistic Regression in Excel

Authored by  
**Mohammed loot**

November 2, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Understanding Confusion Matrices for Logistic Regression in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=8765>

## Introduction to Binary Classification and Model Evaluation

The field of predictive analytics frequently relies on models that can categorize outcomes into one of two states. This process, known as [binary classification](#), is fundamental across diverse disciplines, from finance (predicting loan default) to medicine (diagnosing disease presence). A cornerstone technique for tackling such problems is [Logistic Regression](#), a robust statistical model designed specifically for scenarios where the response variable is **dichotomous** or binary, meaning it can only assume two possible values, typically labeled 0 or 1 (e.g., Yes/No, Success/Failure).

While [Logistic Regression](#) provides a powerful framework for estimating the probability of a specific outcome, its true value is only revealed through rigorous evaluation. It is insufficient merely to train a model; data scientists must employ standardized tools to quantify exactly how well the model's predictions align with the actual, known outcomes of the test data. This critical step ensures the model is reliable and generalizes effectively to new, unseen observations.

The most universally accepted and informative tool for evaluating the performance of any [binary classification](#) model is the [Confusion Matrix](#). This matrix provides an indispensable, granular breakdown of the model's performance, capturing every correct and incorrect prediction, which allows for the calculation of sophisticated performance metrics beyond simple overall accuracy. The following guide details the principles behind the matrix and offers a practical, step-by-step methodology for constructing and utilizing this essential evaluation framework directly within the familiar environment of Microsoft Excel.

		Predicted	
		0	1
Actual	0	30	12
	1	8	56

## Deconstructing the Confusion Matrix: Core Concepts and Components

At its essence, the [Confusion Matrix](#) is a structured 2x2 table that summarizes the relationship between a classification model's predictions and the true, known values (the ground truth). The matrix derives its power from categorizing all outcomes into four distinct categories based on whether the prediction was correct and which class was involved. Understanding these four categories--True Positives, True Negatives, False Positives, and False Negatives--is the absolute foundation of model evaluation.

Before proceeding to the mechanics of calculation in Excel, we must internalize the precise definition of each component, as these definitions will dictate the logical criteria we apply when counting observations in our dataset. We typically designate one outcome as the 'Positive' class (often represented by 1, signifying the event of interest, such as 'fraud' or 'success') and the other as the 'Negative' class (represented by 0).

**True Positives (TP):** These are the successful classifications where the model correctly identified the positive class. In other words, the model predicted 1, and the actual value was also 1. This represents a desirable outcome, indicating the model successfully detected the event of interest.

**True Negatives (TN):** These classifications are also correct, but pertain to the negative class. The model correctly predicted 0, and the actual value was 0. This shows the model effectively ruled out the event of interest when it was not present.

**False Positives (FP):** This is an error, specifically a Type I error. The model incorrectly predicted the positive class (1) when the actual value was negative (0). This is often referred to as a "false alarm" and can carry significant costs depending on the application (e.g., falsely flagging a healthy patient as having a disease).

**False Negatives (FN):** This is the second type of error, a Type II error. The model incorrectly predicted the negative class (0) when the actual value was positive (1). This is a "miss" or a failure to detect the event of interest when it was truly present. High False Negatives are particularly problematic in critical applications like security or medical screening.

Our primary objective in Excel will be to systematically count the total occurrences for each of these four paired outcomes (Actual vs. Predicted) within our test dataset to populate the 2x2 structure of the [Confusion Matrix](#) accurately.

## Practical Setup: Preparing Your Classification Data in Microsoft Excel

The first practical step in constructing the matrix involves ensuring your classification results are properly organized within the spreadsheet. A clean, structured dataset is paramount, requiring at minimum two parallel columns of data that represent the outcomes of your evaluation sample. Specifically, we require the **Actual Values** (the verified ground truth from your held-out test set) and the **Predicted Values** (the outcomes generated by your trained [Logistic Regression](#) model or other classifier).

For clarity and ease of calculation, it is conventional that both the Actual and Predicted values are represented using the binary scheme (0 and 1). Consistency is key: if '1' represents the positive outcome (e.g., a customer will purchase a product), then '0' must represent the negative outcome (the customer will not purchase). The entire process of building the matrix hinges on comparing

these two columns row by row to determine which of the four core outcomes (TP, TN, FP, FN) each observation falls into.

Consider the following example dataset, which contains 20 observations. It is highly recommended to label your columns clearly--for instance, "Actual Class" and "Predicted Class"--to minimize errors when writing the complex conditional counting formulas in the subsequent steps. This setup prepares the necessary ranges for our multi-criteria analysis.

Let's input a sample dataset that includes 20 observations to work with:

	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							

Once the data is correctly structured, define a separate area on your spreadsheet for the [Confusion Matrix](#) itself. This 2x2 area should be clearly labeled with "Actual Class" (0 and 1 along one axis) and "Predicted Class" (0 and 1 along the other axis), creating four empty cells ready to receive the calculated totals for TN, FP, FN, and TP, respectively.

## Building the Matrix: Leveraging the COUNTIFS() Function

The mechanism for constructing the [Confusion Matrix](#) in Excel relies entirely on conditional counting. Unlike simple counting, we need to count rows where two conditions are met

simultaneously (e.g., the Actual value is 1 AND the Predicted value is 0). The native Excel function perfectly suited for this complex task is the **COUNTIFS()** formula. This function enables us to specify multiple criteria and corresponding ranges, ensuring accurate tallying of the four matrix components.

We must write four separate **COUNTIFS()** formulas, one for each cell in the 2x2 matrix. Each formula must reference the two data columns established in Step 1 (Actual and Predicted) and define the specific binary criteria (0 or 1) required for that component. For illustration, assume Column A contains your Actual values and Column B contains your Predicted values.

The four required formulas are structured as follows:

**True Negatives (TN) Calculation:** We are looking for observations where the Actual class is 0 AND the Predicted class is 0. The formula would be: `=COUNTIFS(A:A, 0, B:B, 0)`.

**False Positives (FP) Calculation:** We are looking for errors where the Actual class is 0 BUT the Predicted class is 1. The formula would be: `=COUNTIFS(A:A, 0, B:B, 1)`.

**False Negatives (FN) Calculation:** We are looking for errors where the Actual class is 1 BUT the Predicted class is 0. The formula would be: `=COUNTIFS(A:A, 1, B:B, 0)`.

**True Positives (TP) Calculation:** We are looking for observations where the Actual class is 1 AND the Predicted class is 1. The formula would be: `=COUNTIFS(A:A, 1, B:B, 1)`.

By applying these four precise **COUNTIFS()** formulas to their corresponding cells in the matrix layout, we populate the table with the counts of the four outcomes, thus completing the construction of the raw confusion matrix based on our sample data.

F3								
=COUNTIFS(A2:A21, "=0", B2:B21, "=0")								
	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	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								

	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								

## Deriving Essential Performance Metrics

The raw counts within the [Confusion Matrix](#) are highly informative, but to gain a standardized understanding of the model's quality, we must convert these counts (TP, TN, FP, FN) into measurable metrics. These metrics are expressed as percentages or ratios and are indispensable for comparing the performance of different classification models or different tuning parameters of the same model.

The most fundamental metrics derived from the matrix are [Accuracy](#), [Precision](#), and [Recall](#) (also known as Sensitivity). These ratios are calculated using simple arithmetic operations, referencing the cells of the completed 2x2 matrix.

**[Accuracy](#)**: This metric provides a high-level view of the model's overall correctness. It is defined as the proportion of total correct predictions (both positive and negative) relative to the total number of observations. The formula is:  $(TP + TN) / (TP + TN + FP + FN)$ . While useful, accuracy can be misleading in datasets with severe class imbalance.

**[Precision \(Positive Predictive Value\)](#)**: Precision focuses specifically on the quality of positive

predictions. It addresses the question: "When the model predicts an instance is positive, how likely is that prediction to be correct?" High precision is vital in situations where minimizing False Positives (false alarms) is the priority. The formula is:  $TP / (TP + FP)$ .

**Recall (Sensitivity or True Positive Rate):** Recall focuses on the completeness of positive detection. It addresses the question: "Of all the instances that were truly positive, what fraction did the model successfully identify?" High **Recall** is critical when minimizing False Negatives (missed events) is the priority, such as in identifying fraudulent transactions or medical conditions. The formula is:  $TP / (TP + FN)$ .

To calculate these metrics in Excel, simply set up new cells and insert the formulas, substituting the cell references for your TP, TN, FP, and FN counts. Ensure you format the results as percentages or decimals for clear interpretation.

	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								

### Interpreting the Results and Model Optimization

The final, and arguably most important, stage of the process is the interpretation of the calculated metrics. Referring to the example provided in Step 3, if the calculated **Accuracy** is **0.7** (or 70%), this means that the model correctly predicted the class label (0 or 1) for 70% of the observations in

the test set. While 70% may sound acceptable, further analysis of the [Precision](#) and [Recall](#) scores is essential to truly understand the model's operational profile.

The relationship between [Precision](#) and [Recall](#) often involves a trade-off. A model optimized for extremely high [Recall](#) (e.g., catching nearly all positive cases) may achieve this by being overly eager, resulting in a lower [Precision](#) due to an increased number of False Positives. Conversely, a model optimized for high [Precision](#) (where every positive prediction is highly trustworthy) might miss several true positive cases, leading to lower [Recall](#) and more False Negatives.

This quantitative assessment is vital for the iterative process of model development and selection. By calculating and comparing the [Accuracy](#), [Precision](#), and [Recall](#) of the current [Logistic Regression](#) model against alternative modeling approaches--such as Random Forests, Support Vector Machines, or Neural Networks--we can objectively determine which classifier provides the optimal balance of performance characteristics required for the specific application's risk tolerance and objectives. The [Confusion Matrix](#), therefore, serves not just as an output, but as a diagnostic tool for continuous improvement.

## Next Steps in Classification Model Analysis

While Accuracy, Precision, and Recall offer a strong starting point for model evaluation, the [Confusion Matrix](#) can be leveraged to calculate several other advanced metrics necessary for a complete evaluation, especially when dealing with imbalanced datasets.

To further deepen your understanding of the performance and robustness of [binary classification](#) models, consider exploring methods that combine the core metrics derived from the matrix. These advanced techniques provide alternative views on model efficacy and are often required in professional data science reporting.

**F1 Score Calculation in Excel:** This metric is the harmonic mean of Precision and Recall, offering a balanced measure of performance.

**ROC Curves and AUC:** These tools evaluate model performance across all possible classification thresholds, providing insight into the trade-off between sensitivity and specificity.

**Bias-Variance Tradeoff in Classification Models:** Understanding this concept is crucial for diagnosing whether poor performance is due to underfitting (high bias) or overfitting (high variance).