How to Run Logistic Regression in Excel: A Step-by-Step Guide

Authored by
stats writer

March 12, 2026

Fundamental Concepts of Logistic Regression in Excel

Logistic regression is a sophisticated **statistical method** primarily utilized to model the **probability** of a discrete, **binary outcome** based on one or more **independent variables**. Unlike **linear regression**, which predicts a continuous numerical value, **logistic regression** is designed for scenarios where the **dependent variable** is categorical, typically representing two possible states such as "success" versus "failure," "yes" versus "no," or "drafted" versus "not drafted." This makes it an indispensable tool in fields ranging from medicine and social sciences to sports analytics and business forecasting.

While specialized statistical software like **R**, **SPSS**, or **Python**'s **Scikit-Learn** library are often preferred for complex modeling, **Microsoft Excel** remains a highly accessible and powerful platform for performing these analyses. Utilizing **Excel** for **logistic regression** allows users to leverage familiar spreadsheet functionalities while implementing **Maximum Likelihood Estimation** (MLE) to determine the best-fitting model parameters. By organizing raw data systematically and employing optimization tools, analysts can derive meaningful insights without the need for advanced programming knowledge.

The core objective of **logistic regression** is to describe the relationship between a **binary** response variable and several explanatory variables by fitting a **logistic function**. This function produces an S-shaped curve that constrains the predicted values between 0 and 1, ensuring that the results represent valid **probabilities**. Throughout this tutorial, we will explore the step-by-step process of configuring your data, defining the mathematical components of the model, and utilizing **Excel**'s optimization features to generate a predictive equation.

In the context of data science, the **logit** transformation is essential because it maps **probability** values to the entire range of real numbers, allowing for the application of linear modeling techniques. This tutorial will focus on a practical application involving college basketball players, providing a clear roadmap for anyone looking to perform rigorous **data analysis** within a spreadsheet environment. By the end of this guide, you will be equipped to handle **binary classification** problems and interpret the results with high precision.

Step 1: Systematic Organization of Raw Data

The initial phase of any **regression analysis** involves meticulous **data preparation**. In this example, we are examining a dataset that tracks whether college basketball players were drafted into the **NBA** based on three specific performance metrics: average points per game (pts), average rebounds per game (rebs), and average assists per game (ast). The **dependent variable**, "draft," is encoded as a **binary** indicator where 0 signifies the player was not drafted and 1 signifies they were successfully drafted.

Proper formatting is crucial for the formulas to function correctly across multiple rows. You should arrange your **independent variables** in adjacent columns, followed by the outcome column. This structured layout ensures that the **matrix multiplication** or linear combinations required for the model can be calculated efficiently. Consistency in data entry prevents errors during the optimization phase, where **Excel** will iterate through values to find the most accurate coefficients.

Ensure that there are no missing values in your dataset, as **logistic regression** models are sensitive to incomplete data points. In **Excel**, you can use the "Filter" or "Sort" functions to identify and address any anomalies before proceeding. Once your data is clean and correctly labeled, you can proceed to the next stage of setting up the model's structural components.

	A	B	C	D	E	F
1	**draft**	**pts**	**rebs**	**ast**		
2	0	12	3	6		
3	1	13	4	4		
4	0	13	4	6		
5	1	12	9	9		
6	1	14	4	5		
7	0	14	4	4		
8	0	17	2	2		
9	1	17	6	5		
10	1	21	5	7		
11	0	21	9	3		
12	1	24	11	11		
13	0	24	4	5		
14						
15						
16						
17						
18						
19						
20						
21						
22						

Step 2: Initializing Regression Coefficients and Placeholders

To begin the modeling process, you must establish a dedicated area for the **regression coefficients**. Since our model includes three explanatory variables (points, rebounds, and assists), we require three specific coefficients plus one additional value for the **y-intercept**. The **intercept** represents the baseline **logit** value when all **independent variables** are equal to zero, serving as the starting point for the model's predictions.

Initially, we populate these coefficient cells with a small placeholder value, such as **0.001**. These values are temporary; they serve as the starting point for the **Solver** algorithm, which will eventually optimize them to maximize the **likelihood** of the observed data. Placing these coefficients in a separate, clearly labeled section of your spreadsheet--such as cells B15 through B18--makes it easier to reference them in the upcoming formulas using **absolute cell references**.

In **statistical modeling**, these coefficients determine the weight and direction of the relationship between each predictor and the outcome. A positive coefficient suggests that an increase in the predictor increases the **probability** of the event occurring, while a negative coefficient indicates the opposite. Setting these initial values is a prerequisite for the iterative **maximum likelihood** process that characterizes **logistic regression**.

	A	B	C	D	E	F
1	**draft**	**pts**	**rebs**	**ast**		
2	0	12	3	6		
3	1	13	4	4		
4	0	13	4	6		
5	1	12	9	9		
6	1	14	4	5		
7	0	14	4	4		
8	0	17	2	2		
9	1	17	6	5		
10	1	21	5	7		
11	0	21	9	3		
12	1	24	11	11		
13	0	24	4	5		
14						
15	**b0**	0.001				
16	**b1**	0.001				
17	**b2**	0.001				
18	**b3**	0.001				
19						
20						
21						
22						

Following the initialization of the coefficients, you must prepare additional columns to facilitate the transformation of raw data into **probabilistic** estimates. These columns will include the **logit** (the linear combination of predictors and coefficients), the **exponentiated logit** (e^logit), the predicted **probability**, and the **log likelihood**. This stepwise approach allows for a transparent view of the model's internal calculations.

Step 3: Calculating Logit Values and Linear Combinations

The **logit** represents the linear predictor part of the **logistic regression** equation. It is calculated by multiplying each **independent variable** by its corresponding coefficient and adding the **intercept**. In **Excel**, this is typically achieved using a formula that references both the data in the current row and the fixed coefficient cells established in the previous step. Using **absolute references** (e.g., $B$15) is vital here to ensure the formula correctly identifies the coefficients as it is dragged down through the dataset.

The mathematical expression for the **logit** is: **Logit = Intercept + (C1 * Var1) + (C2 * Var2) + (C3 * Var3)**. This value is technically the log-odds of the dependent variable being equal to 1. Because the log-odds can range from negative infinity to positive infinity, it provides a continuous scale that can be mapped back to a 0-1 **probability** range using the **logistic function**. Calculating this for every observation in your dataset is the first step toward generating individualized predictions.

By viewing the **logit** column, you can see how different combinations of player stats contribute to the overall score. At this stage, because we used 0.001 as a placeholder, the logit values will appear very similar across all rows. This will change dramatically once the **Solver** is executed and the coefficients are optimized to reflect the actual patterns found in the **NBA** draft data.

	A	B	C	D	E	F	G	H	I
1	**draft**	**pts**	**rebs**	**ast**	**Logit**				
2	0	12	3	6	0.022	=$B$15+$B$16*B2 + $B$17*C2 + $B$18*D2			
3	1	13	4	4	0.022				
4	0	13	4	6	0.024				
5	1	12	9	9	0.031				
6	1	14	4	5	0.024				
7	0	14	4	4	0.023				
8	0	17	2	2	0.022				
9	1	17	6	5	0.029				
10	1	21	5	7	0.034				
11	0	21	9	3	0.034				
12	1	24	11	11	0.047				
13	0	24	4	5	0.034				
14									
15	**b0**	0.001							
16	**b1**	0.001							
17	**b2**	0.001							
18	**b3**	0.001							
19									
20									
21									
22									

Step 4: Transforming Logits into Probabilities

Once the **logit** values are established, the next requirement is to transform these values into **probabilities**. This involves two distinct mathematical steps. First, we calculate **e^logit**, which is the natural logarithm base (approximately 2.718) raised to the power of the logit value. In **Excel**, the **EXP()** function is utilized for this purpose. This transformation is a critical component of the **sigmoid function**, which is the hallmark of **logistic regression**.

	A	B	C	D	E	F	G
1	draft	pts	rebs	ast	Logit	e^{Logit}	
2	0	12	3	6	0.022	1.0222	=EXP(E2)
3	1	13	4	4	0.022	1.0222	
4	0	13	4	6	0.024	1.0243	
5	1	12	9	9	0.031	1.0315	
6	1	14	4	5	0.024	1.0243	
7	0	14	4	4	0.023	1.0233	
8	0	17	2	2	0.022	1.0222	
9	1	17	6	5	0.029	1.0294	
10	1	21	5	7	0.034	1.0346	
11	0	21	9	3	0.034	1.0346	
12	1	24	11	11	0.047	1.0481	
13	0	24	4	5	0.034	1.0346	
14							
15	b0	0.001					
16	b1	0.001					
17	b2	0.001					
18	b3	0.001					
19							
20							
21							

After calculating **e^logit**, we determine the actual **probability** using the formula: **Probability = e^logit / (1 + e^logit)**. This calculation ensures that every predicted value falls strictly between 0 and 1, representing the likelihood that a specific player will be drafted. This **probability** is the core output of the model, allowing analysts to set thresholds (commonly 0.5) for classification purposes.

The relationship between the **logit** and the **probability** is non-linear. Small changes in the **independent variables** have the greatest impact on the **probability** when the initial **probability** is near 0.5, and less impact as it approaches 0 or 1. This characteristic makes **logistic regression** particularly robust for modeling outcomes where the effect of predictors levels off at extreme values. Understanding this transformation is key to interpreting how the player's stats influence their draft prospects.

SUM =IF(A2=1, F2/(1+F2), 1-(F2/(1+F2)))

	A	B	C	D	E	F	G	H	I	J
1	**draft**	**pts**	**rebs**	**ast**	**Logit**	**e^{Logit}**	**probability**			
2	0	12	3	6	0.022	1.0222	0.4945002	=IF(A2=1, F2/(1+F2), 1-(F2/(1+F2)))		
3	1	13	4	4	0.022	1.0222	0.5054998			
4	0	13	4	6	0.024	1.0243	0.4940003			
5	1	12	9	9	0.031	1.0315	0.5077494			
6	1	14	4	5	0.024	1.0243	0.5059997			
7	0	14	4	4	0.023	1.0233	0.4942503			
8	0	17	2	2	0.022	1.0222	0.4945002			
9	1	17	6	5	0.029	1.0294	0.5072495			
10	1	21	5	7	0.034	1.0346	0.5084992			
11	0	21	9	3	0.034	1.0346	0.4915008			
12	1	24	11	11	0.047	1.0481	0.5117478			
13	0	24	4	5	0.034	1.0346	0.4915008			
14										
15	**b0**	0.001								
16	**b1**	0.001								
17	**b2**	0.001								
18	**b3**	0.001								
19										
20										

Step 5: Implementing the Log Likelihood Function

To determine the accuracy of our model, we must use a metric called the **log likelihood**. This value represents the **logarithm** of the **probability** that the model would produce the observed outcomes given the current coefficients. For **logistic regression**, we calculate the **log likelihood** for each individual observation. The goal of the optimization process is to find coefficient values that make the observed data as "likely" as possible.

The formula for **log likelihood** in this context is **LN(Probability)**. However, in a full **Maximum Likelihood Estimation**, the formula typically accounts for both the "success" and "failure" cases. For the purposes of this **Excel** tutorial, we focus on the log of the predicted **probability** for the actual observed class. By summing these individual log likelihoods, we arrive at a single value that characterizes the overall "fit" of the model. This sum is the "Objective" that we will ask **Excel** to maximize.

It is important to note that **log likelihood** values are typically negative because **probabilities** are less than 1, and the natural log of a fraction is negative. Therefore, when we "maximize" the **log likelihood**, we are actually moving the value closer to zero. A sum of **log likelihoods** that is closer to zero indicates a model that predicts the observed outcomes with much higher confidence and accuracy.

	A	B	C	D	E	F	G	H	I
1	**draft**	**pts**	**rebs**	**ast**	**Logit**	**e^{Logit}**	**probability**	**log likelihood**	
2	0	12	3	6	0.022	1.0222	0.4945002	-0.704207679	=LN(G2)
3	1	13	4	4	0.022	1.0222	0.5054998	-0.682207679	
4	0	13	4	6	0.024	1.0243	0.4940003	-0.705219179	
5	1	12	9	9	0.031	1.0315	0.5077494	-0.677767301	
6	1	14	4	5	0.024	1.0243	0.5059997	-0.681219179	
7	0	14	4	4	0.023	1.0233	0.4942503	-0.704713304	
8	0	17	2	2	0.022	1.0222	0.4945002	-0.704207679	
9	1	17	6	5	0.029	1.0294	0.5072495	-0.678752302	
10	1	21	5	7	0.034	1.0346	0.5084992	-0.676291674	
11	0	21	9	3	0.034	1.0346	0.4915008	-0.710291674	
12	1	24	11	11	0.047	1.0481	0.5117478	-0.66992328	
13	0	24	4	5	0.034	1.0346	0.4915008	-0.710291674	
14									
15	**b0**	0.001							
16	**b1**	0.001							
17	**b2**	0.001							
18	**b3**	0.001							
19									
20									

	A	B	C	D	E	F	G	H	I
1	**draft**	**pts**	**rebs**	**ast**	**Logit**	**e^{Logit}**	**probability**	**log likelihood**	
2	0	12	3	6	0.022	1.0222	0.4945002	-0.704207679	
3	1	13	4	4	0.022	1.0222	0.5054998	-0.682207679	
4	0	13	4	6	0.024	1.0243	0.4940003	-0.705219179	
5	1	12	9	9	0.031	1.0315	0.5077494	-0.677767301	
6	1	14	4	5	0.024	1.0243	0.5059997	-0.681219179	
7	0	14	4	4	0.023	1.0233	0.4942503	-0.704713304	
8	0	17	2	2	0.022	1.0222	0.4945002	-0.704207679	
9	1	17	6	5	0.029	1.0294	0.5072495	-0.678752302	
10	1	21	5	7	0.034	1.0346	0.5084992	-0.676291674	
11	0	21	9	3	0.034	1.0346	0.4915008	-0.710291674	
12	1	24	11	11	0.047	1.0481	0.5117478	-0.66992328	
13	0	24	4	5	0.034	1.0346	0.4915008	-0.710291674	
14								-8.305092603	=SUM(H2:H13)
15	**b0**	0.001							
16	**b1**	0.001							
17	**b2**	0.001							
18	**b3**	0.001							
19									
20									
21									

Step 6: Activating and Configuring the Excel Solver

The **Solver** is a powerful **Excel Add-in** used for "what-if" analysis and **optimization**. It is not always enabled by default, so you may need to install it manually. Navigate to the **File** tab, select **Options**, and then click on **Add-Ins**. In the "Manage" dropdown at the bottom, ensure "Excel Add-ins" is selected and click **Go**. Check the box next to **Solver Add-In** and click **OK**. Once installed, the **Solver** tool will appear in the **Analysis** group on the **Data** tab.

To configure the **Solver** for **logistic regression**, you must define the following parameters in the dialog box:

Set Objective: Select the cell containing the sum of your **log likelihoods** (e.g., H14).
To: Select the **Max** radio button, as we want to maximize the **likelihood**.
By Changing Variable Cells: Select the range of cells containing your initial coefficients (e.g., B15:B18).
Make Unconstrained Variables Non-Negative: This box must be **unchecked** because **regression coefficients** can be either positive or negative.
Select a Solving Method: Choose **GRG Nonlinear**, which is designed for optimization problems where the relationship between variables is not a straight line.

After entering these details, click **Solve**. **Excel** will then perform a series of iterations, adjusting the coefficients until it finds the combination that maximizes the **log likelihood**. If successful, a window will appear stating that **Solver** found a solution. Choose to keep the **Solver** solution to update your coefficient cells with the optimized values.

Step 7: Interpreting Results and Coefficient Adjustment

Once the **Solver** completes its task, the cells that previously held 0.001 will now contain the calculated **regression coefficients**. However, there is a nuance in how **Excel**'s optimization might categorize the output. By default, the optimization we performed often calculates the **probability** that the response variable is 0. In many **statistical** contexts, and specifically in this tutorial, we are more interested in the **probability** that the response variable equals 1 (i.e., the player is drafted).

	A	B	C	D	E	F	G	H
1	**draft**	**pts**	**rebs**	**ast**	**Logit**	**e^{Logit}**	**probability**	**log likelihood**
2	0	12	3	6	0.229211	1.2576	0.4429467	-0.814305767
3	1	13	4	4	-0.84702	0.4287	0.3000576	-1.203780974
4	0	13	4	6	0.512049	1.6687	0.3747132	-0.981594287
5	1	12	9	9	4.641846	103.7357	0.9904522	-0.009593714
6	1	14	4	5	-0.28032	0.7555	0.4303753	-0.843097633
7	0	14	4	4	-0.95986	0.3829	0.7230931	-0.32421729
8	0	17	2	2	-3.44877	0.0318	0.9691944	-0.031290023
9	1	17	6	5	0.172523	1.1883	0.5430241	-0.610601525
10	1	21	5	7	0.684594	1.9830	0.6647633	-0.408324279
11	0	21	9	3	-0.45087	0.6371	0.6108458	-0.492910763
12	1	24	11	11	5.438267	230.0432	0.9956718	-0.004337588
13	0	24	4	5	-1.40865	0.2445	0.8035527	-0.218712509
14								-5.942766351
15	**b0**	-3.68103						
16	**b1**	-0.11283						
17	**b2**	0.395671						
18	**b3**	0.679537						
19								
20								

To align the results with the **probability** of a "success" (Draft = 1), we simply reverse the signs of each coefficient. For instance, if the **Solver** returned a positive coefficient for points, we change it to negative, and vice versa. This mathematical adjustment ensures that the final model correctly predicts the **likelihood** of being drafted based on the performance metrics provided. This step is essential for maintaining the **interpretability** of the model.

	A	B	C	D	E	F	G	H
1	draft	pts	rebs	ast	Logit	e^{Logit}	probability	log likelihood
2	0	12	3	6	0.229211	1.2576	0.4429467	-0.814305767
3	1	13	4	4	-0.84702	0.4287	0.3000576	-1.203780974
4	0	13	4	6	0.512049	1.6687	0.3747132	-0.981594287
5	1	12	9	9	4.641846	103.7357	0.9904522	-0.009593714
6	1	14	4	5	-0.28032	0.7555	0.4303753	-0.843097633
7	0	14	4	4	-0.95986	0.3829	0.7230931	-0.32421729
8	0	17	2	2	-3.44877	0.0318	0.9691944	-0.031290023
9	1	17	6	5	0.172523	1.1883	0.5430241	-0.610601525
10	1	21	5	7	0.684594	1.9830	0.6647633	-0.408324279
11	0	21	9	3	-0.45087	0.6371	0.6108458	-0.492910763
12	1	24	11	11	5.438267	230.0432	0.9956718	-0.004337588
13	0	24	4	5	-1.40865	0.2445	0.8035527	-0.218712509
14		P(X=0)	P(X=1)					-5.942766351
15	b0	-3.68103	3.68103					
16	b1	-0.11283	0.11283					
17	b2	0.395671	-0.395671					
18	b3	0.679537	-0.679537					
19								
20								
21								
22								

The finalized coefficients represent the **log-odds** impact of each variable. For example, a positive coefficient for points implies that as a player's points per game increase, their **probability** of being drafted also increases. Conversely, if a variable like "turnovers" (not included here, but possible in other models) had a negative coefficient, it would mean that higher turnovers decrease the draft **probability**. These insights allow for a nuanced understanding of the factors driving **NBA** draft decisions.

Step 8: Applying the Model for Real-World Predictions

With the optimized coefficients in hand, you can now use the **logistic regression** equation to predict outcomes for new data. Consider a hypothetical college player who averages 14 points, 4 rebounds, and 5 assists. To find the **probability** of this player being drafted, we plug these values into our equation using the adjusted coefficients. The formula follows the standard **logistic** structure: **P(Draft=1) = e^(Linear Combination) / (1 + e^(Linear Combination))**.

Using the specific coefficients derived in our example, the calculation would look like this: **P(draft = 1) = e^(3.681 + 0.113*14 - 0.396*4 - 0.680*5) / (1 + e^(3.681 + 0.113*14 - 0.396*4 - 0.680*5))**. Performing this math results in a **probability** of **0.57**. In **statistical classification**, a common threshold is 0.5; since 0.57 is greater than this cutoff, we would formally predict that this player is likely to be drafted into the **NBA**.

This predictive capability is the primary value of **regression analysis**. Beyond simply describing historical data, the model serves as a functional tool for **decision-making**. Scouts, coaches, and analysts can use these formulas to evaluate prospects objectively. Furthermore, by adjusting the input variables, one can perform **sensitivity analysis** to see how much a player would need to improve their rebounding or scoring to significantly boost their draft **probability**.

It is important to remember that while **Excel** is an excellent tool for learning and performing basic **logistic regression**, it does not automatically provide some of the diagnostic statistics found in specialized software, such as **p-values** for individual coefficients or **R-squared** equivalents like the **Hosmer-Lemeshow test**. However, for many practical applications and data exploration tasks, the method outlined here provides a robust and reliable way to implement **predictive modeling** and **data analysis**.