

# Logistic Regression Tutorial: A Step-by-Step Guide with Excel Examples

Authored by  
**Mohammed Iooti**

November 8, 2025

## RECOMMENDED CITATION

Mohammed Iooti (2025). *Logistic Regression Tutorial: A Step-by-Step Guide with Excel Examples*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13190>

**Logistic regression** is a foundational statistical method employed when the predicted outcome or **response variable** is **binary** (dichotomous). Unlike traditional linear regression, which forecasts a continuous numerical result, logistic regression models the inherent probability of a specific event occurring. This methodology is indispensable across various high-stakes fields, including financial risk assessment, medical diagnostics, and sophisticated predictive analytics in sports.

While professional statistical software packages (such as R or SPSS) are typically used for large-scale analysis, **Microsoft Excel** possesses the robust numerical capabilities necessary to execute complex **logistic regression** calculations manually. By leveraging the built-in Solver Add-In, we can perform the iterative optimization required to derive accurate model parameters. This comprehensive, step-by-step tutorial demystifies the process of conducting a complete logistic regression analysis directly within the Excel environment, ensuring both accuracy and deep interpretability of the final results.

## Defining the Model: A Sports Analytics Case Study

To demonstrate the practical application of this technique, we will utilize a dataset focused on college basketball players. Our objective is to construct a predictive model that determines whether a player is drafted into the NBA. The dependent **binary outcome** is defined as Draft status (1 = drafted; 0 = not drafted).

The prediction will be based on three key independent explanatory variables derived from the players' previous season performance: average points (pts), average rebounds (rebs), and average assists (ast). The goal is to isolate the specific influence of each performance metric on the likelihood of achieving professional status.

## Structuring Data and Initial Parameter Setup

The initial step requires meticulous organization of the raw data within the Excel spreadsheet. It is vital to structure the input clearly, separating the dependent variable (Draft status) from the independent explanatory variables (pts, rebs, ast) into distinct, labeled columns. Accuracy in this data entry phase is critical, as any error will propagate through the complex optimization routine.

	A	B	C	D	E	F
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>		
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						

Since our predictive model incorporates three distinct explanatory variables, we require a total of four **regression coefficients** for the analysis: one coefficient corresponding to each variable and an additional coefficient representing the model's **intercept**. Designate a specific range of cells (e.g., B15:B18) to hold these coefficients. Initially, assign a small, arbitrary non-zero value (such as 0.001) to each cell. These placeholder values are essential starting points that the Excel Solver will iteratively adjust to optimize the model's fit.

	A	B	C	D	E	F
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>		
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	<b>b0</b>	0.001				
16	<b>b1</b>	0.001				
17	<b>b2</b>	0.001				
18	<b>b3</b>	0.001				
19						
20						
21						
22						

Following the initialization of the coefficients, we must prepare the spreadsheet for the calculation sequence. This involves establishing several new columns that are fundamental to the iterative process of Maximum Likelihood Estimation. These new calculations include the [logit](#) (the linear predictor), the exponential transformation of the logit (elogit), the resulting predicted probability (P), and finally, the [log likelihood](#) score.

## Modeling the Linear Predictor and Probability

The first major calculation is the [logit](#) function. Conceptually, the logit is the linear core of the logistic regression model, representing the weighted sum of the predictor variables and their corresponding initial [regression coefficients](#). This calculation transforms the binary prediction problem into a linear scale. We generate the logit column by applying the following formula to each observation in the dataset:  $\text{Logit} = \text{Intercept} + (\text{Coefficient}_{\text{pts}} * \text{pts}) + (\text{Coefficient}_{\text{rebs}} * \text{rebs}) + (\text{Coefficient}_{\text{ast}} * \text{ast})$ .

	A	B	C	D	E	F	G	H	I
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>	<b>Logit</b>				
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	<b>b0</b>	0.001							
16	<b>b1</b>	0.001							
17	<b>b2</b>	0.001							
18	<b>b3</b>	0.001							
19									
20									
21									
22									

The logit value, which can range from negative to positive infinity, must be converted into a probability value that is constrained between 0 and 1. The first step in this conversion is calculating the exponential of the **logit** value (elogit). This is easily accomplished in Excel using the built-in `EXP()` function, linking it directly to the cell containing the calculated logit for that observation.

	A	B	C	D	E	F	G
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>	<b>Logit</b>	<b>e<sup>Logit</sup></b>	
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	<b>b0</b>	0.001					
16	<b>b1</b>	0.001					
17	<b>b2</b>	0.001					
18	<b>b3</b>	0.001					
19							
20							
21							

With elogit established, we can now calculate the final predicted probability (P) that the event occurs (i.e., Draft = 1). The standard logistic function mathematically defines this probability as  $P = \text{elogit} / (1 + \text{elogit})$ . This formula is implemented across the entire probability column, providing an initial probability estimate for every player based on the arbitrary starting coefficient values.

	A	B	C	D	E	F	G	H	I	J
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>	<b>Logit</b>	<b>e<sup>Logit</sup></b>	<b>probability</b>			
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	<b>b0</b>	0.001								
16	<b>b1</b>	0.001								
17	<b>b2</b>	0.001								
18	<b>b3</b>	0.001								
19										
20										

### Implementing the Maximum Likelihood Criterion

In [logistic regression](#), the model is calibrated by maximizing the [log likelihood](#) function. This function serves as the key metric of model fitness, quantifying how accurately the predicted probabilities align with the actual observed binary outcomes (0 or 1). For each observation, the log likelihood is calculated using the natural logarithm (LN) of the resulting probability (P) or its complement (1-P), depending on the outcome (Y):

$$\text{Log likelihood} = \text{LN}(\text{Probability})$$

Populate the log likelihood column using this natural logarithm calculation, referencing the probability value derived in the previous step.

	A	B	C	D	E	F	G	H	I
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>	<b>Logit</b>	<b>e<sup>Logit</sup></b>	<b>probability</b>	<b>log likelihood</b>	
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	<b>b0</b>	0.001							
16	<b>b1</b>	0.001							
17	<b>b2</b>	0.001							
18	<b>b3</b>	0.001							
19									
20									
21									

The overall goal of the optimization process is to determine the parameter estimates (the [regression coefficients](#)) that yield the highest possible fit across the entire dataset. To prepare for optimization, create a dedicated cell (e.g., H14) to calculate the total sum of all individual log likelihood scores. This single, aggregated value is the objective function that the Excel Solver will manipulate and maximize.

	A	B	C	D	E	F	G	H	I
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>	<b>Logit</b>	$e^{\text{Logit}}$	<b>probability</b>	<b>log likelihood</b>	
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	<b>b0</b>	0.001							
16	<b>b1</b>	0.001							
17	<b>b2</b>	0.001							
18	<b>b3</b>	0.001							
19									
20									
21									

## Optimizing Parameters Using the Excel Solver

The **Solver** Add-In is the engine of this analysis, providing the necessary numerical optimization to achieve Maximum Likelihood Estimation. If the Solver icon is not visible in the Analysis group under the Data tab, it must first be installed:

Click **File**.

Click **Options**.

Click **Add-Ins**.

Select the **Solver Add-In**, then click **Go**.

In the new dialog box, check the box next to **Solver Add-In**, then click **OK**.

Once installed, navigate to the **Data** tab and click **Solver** to configure the optimization parameters. Careful configuration is necessary to ensure the calculation converges correctly toward the maximum **log likelihood**:

**Set Objective:** Select the cell containing the calculated sum of the log likelihoods (e.g., H14). Critically, ensure the objective is set to **Max** (Maximize).

**By Changing Variable Cells:** Select the range containing the initial **regression coefficients** (e.g., B15:B18). These are the values the Solver will adjust.

**Make Unconstrained Variables Non-Negative:** This box must be **unchecked**, as regression

coefficients in logistic models frequently take on negative values.

**Select a Solving Method:** Choose **GRG Nonlinear**. This method is specifically designed for complex, non-linear optimization problems like maximum likelihood estimation in [logistic regression](#).

Click **Solve**. The Solver will run its iterations, finding the optimal set of coefficient values that maximize the objective function.

## Interpreting Results and Predictive Application

Upon successful convergence, the Solver will display the final, optimized [regression coefficients](#) that provide the best predictive fit for the basketball drafting data:

	A	B	C	D	E	F	G	H
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>	<b>Logit</b>	$e^{\text{Logit}}$	<b>probability</b>	<b>log likelihood</b>
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	<b>b0</b>	-3.68103						
16	<b>b1</b>	-0.11283						
17	<b>b2</b>	0.395671						
18	<b>b3</b>	0.679537						
19								
20								

A crucial point of interpretation is that the default coefficients derived in this specific setup are designed to predict the probability that the event *does not* occur ( $P(\text{Draft} = 0)$ ). However, conventional statistical reporting and practical application usually focus on predicting the success case ( $P(\text{Draft} = 1)$ ).

To correct this and make the coefficients interpretable for predicting  $P(\text{Draft} = 1)$ , you must simply reverse the sign of every calculated **regression coefficient** (multiplying each by -1).

	A	B	C	D	E	F	G	H
1	<b>draft</b>	<b>pts</b>	<b>rebs</b>	<b>ast</b>	<b>Logit</b>	<b>e<sup>Logit</sup></b>	<b>probability</b>	<b>log likelihood</b>
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	<b>b0</b>	-3.68103	3.68103					
16	<b>b1</b>	-0.11283	0.11283					
17	<b>b2</b>	0.395671	-0.395671					
18	<b>b3</b>	0.679537	-0.679537					
19								
20								
21								
22								

These adjusted coefficients are now ready for practical predictive use. Consider a new hypothetical player who averaged 14 points, 4 rebounds, and 5 assists. We can calculate the probability of this player being drafted ( $P(\text{Draft} = 1)$ ) by plugging these values into the full logistic function equation:

$$P(\text{draft} = 1) = \frac{e^{3.681193 + 0.112827*(14) - 0.39568*(4) - 0.67954*(5)}}{1 + e^{3.681193 + 0.112827*(14) - 0.39568*(4) - 0.67954*(5)}} = \mathbf{0.57}.$$

Since the resulting probability (0.57) exceeds the standard classification threshold of 0.5, our [logistic regression](#) model successfully predicts that this hypothetical player has a high likelihood of being drafted into the NBA. This exercise confirms that Excel, when combined with the Solver Add-In, is a powerful and accessible platform for executing and applying advanced statistical modeling techniques.

#### Related Resources: