

# Learning Logistic Regression with SPSS: A Step-by-Step Tutorial

Authored by  
**Mohammed loot**

November 8, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Learning Logistic Regression with SPSS: A Step-by-Step Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12873>

The field of statistical modeling often requires techniques capable of handling outcomes that are inherently categorical rather than continuous. [Logistic regression](#) is a powerful method used extensively across disciplines to fit a [regression model](#) specifically when the response variable is dichotomous or **binary** (e.g., Yes/No, Success/Failure, Drafted/Not Drafted).

This comprehensive tutorial provides a detailed, step-by-step guide on how to successfully perform and interpret [logistic regression](#) using [SPSS](#) (Statistical Package for the Social Sciences), ensuring clarity and accuracy in your analysis.

## Understanding Logistic Regression and its Applications

Unlike standard linear regression, which predicts a continuous outcome, [logistic regression](#) models the probability of a particular event occurring. It transforms the linear relationship using the logit function, ensuring that the predicted probabilities fall strictly between 0 and 1. This mathematical foundation makes it indispensable for analyzing situations where the outcome must belong to one of two distinct categories.

The primary goal of this analysis is twofold: first, to determine which predictor variables significantly influence the binary outcome; and second, to calculate the probability of the event occurring based on specific values of the predictors. Understanding these probabilities is crucial for forecasting events, risk assessment, and decision-making in fields ranging from medicine (disease presence) to finance (loan default) and, as we will explore, sports analytics (player drafting).

In the context of this tutorial, we will be examining how collegiate performance metrics relate to professional outcomes. We aim to establish a robust model within the [SPSS](#) environment that can accurately predict a college basketball player's likelihood of being drafted into the NBA based on quantifiable performance data.

## Setting Up the Analysis in SPSS (The Case Study)

To demonstrate the mechanics of the procedure, we will use a hypothetical yet realistic scenario. We have compiled a dataset tracking college basketball players, recording whether or not they were drafted into the **NBA**. This outcome variable, **draft**, is a [binary response variable](#) (coded as 0 for 'not drafted' and 1 for 'drafted').

The analysis seeks to determine how two specific [predictor variables](#) influence this outcome: the player's average points per game (a continuous variable) and the division level of their college program (a categorical variable). This setup allows us to explore how both individual performance metrics and institutional context contribute to professional selection.

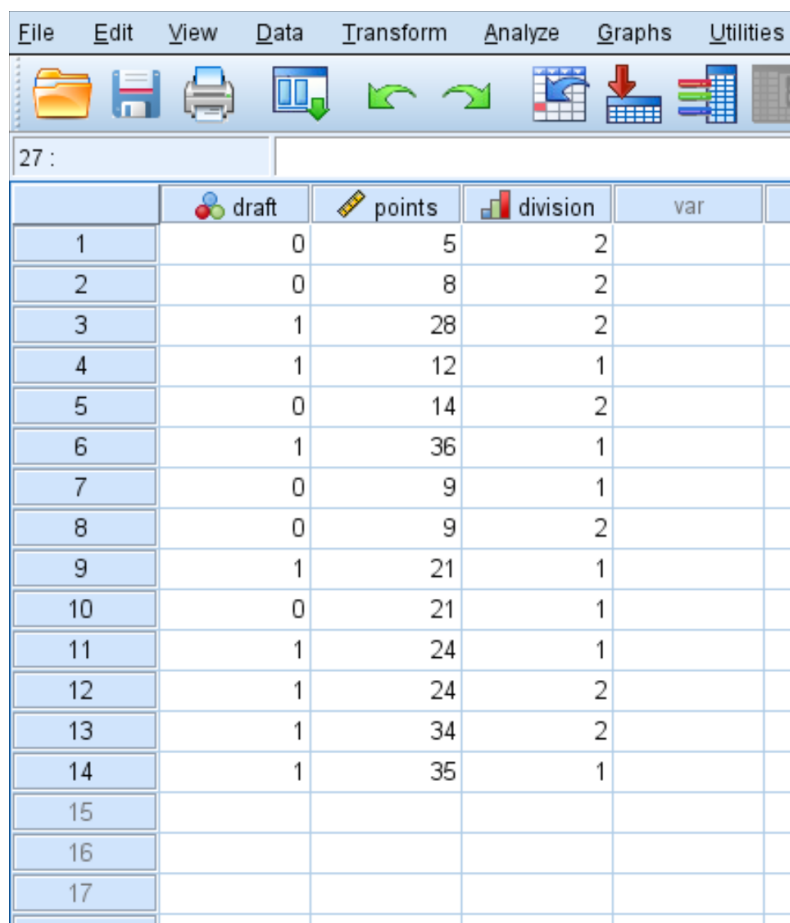
Before proceeding with the regression, it is imperative that the data is correctly entered and

structured within the [SPSS](#) Data View and Variable View windows. Proper data management ensures that the software correctly recognizes the scale and type of each variable, which is critical for accurate model execution.

## Step-by-Step Data Preparation and Input

The initial and most critical step involves accurately inputting the raw data into [SPSS](#). Our dataset consists of 14 observations, capturing the three key variables: the binary outcome (draft), and the two predictors (points and division). Ensure that the Variable View defines 'draft' appropriately as a nominal or scale variable depending on your software version, but confirming its binary coding (0=no, 1=yes) is essential.

The following table illustrates the structure of the data as it should appear in the [SPSS](#) environment:



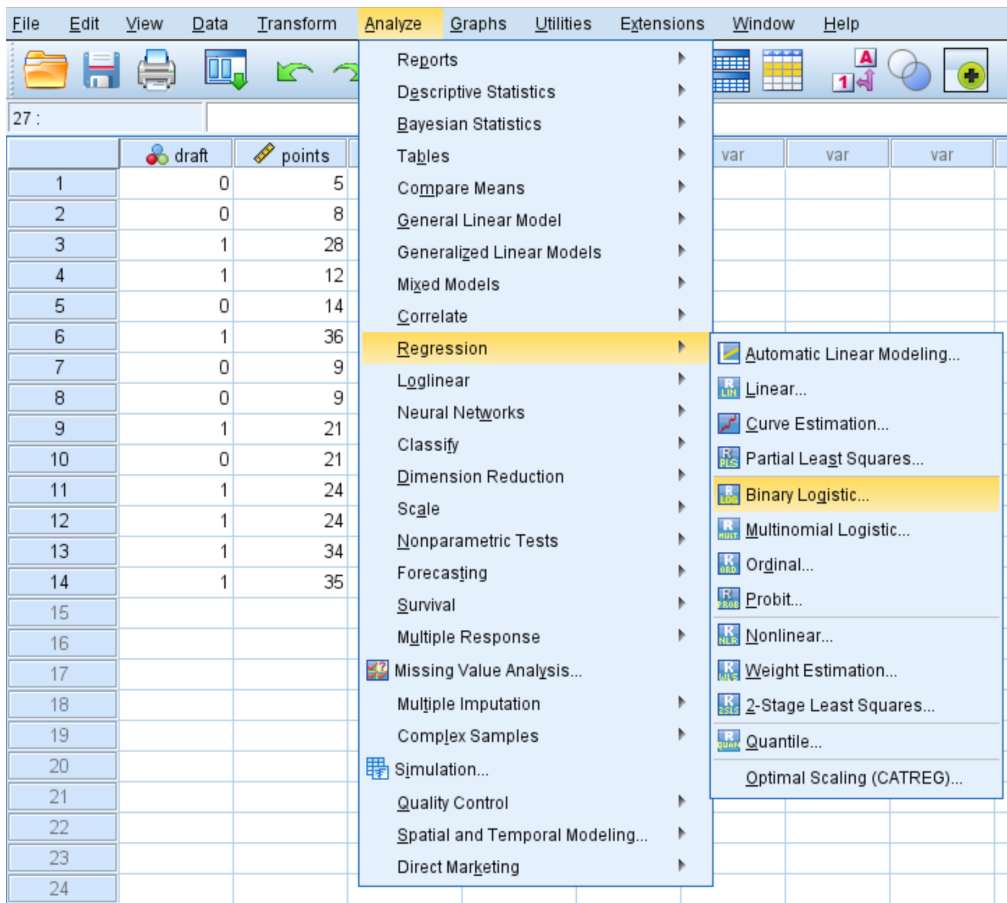
The screenshot shows the SPSS Data View window with a menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities) and a toolbar. Below the toolbar, the variable names are listed: draft, points, division, and var. The data is organized into 14 rows, each representing an observation. The 'draft' column contains binary values (0 or 1), 'points' contains numerical values, and 'division' contains numerical values (1 or 2). The 'var' column is empty.

	draft	points	division	var
1	0	5	2	
2	0	8	2	
3	1	28	2	
4	1	12	1	
5	0	14	2	
6	1	36	1	
7	0	9	1	
8	0	9	2	
9	1	21	1	
10	0	21	1	
11	1	24	1	
12	1	24	2	
13	1	34	2	
14	1	35	1	
15				
16				
17				
18				

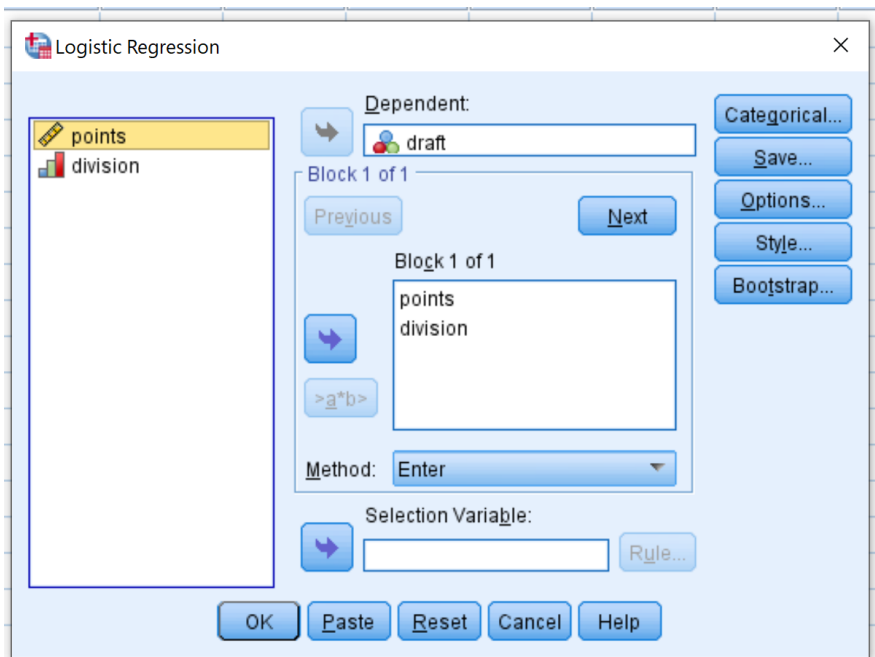
Once the data is verified, we can proceed to the analysis stage. The integrity of the final model relies heavily on the correctness of this initial input stage, particularly ensuring that the **binary response variable** is correctly coded for the desired outcome (1 = success).

## Executing the Binary Logistic Regression Procedure

With the data ready, the next step is to initiate the logistic regression procedure through the graphical user interface of **SPSS**. Navigate to the menu bar and click the **Analyze** tab. From the dropdown menu, hover over **Regression**, and then select the option for **Binary Logistic Regression**. This selection opens the primary dialogue box where the model specification takes place.



In the dialogue box that appears, you must correctly assign the variables to their respective roles. Drag the **binary response variable**, **draft**, into the box labeled **Dependent**. Subsequently, move the two **predictor variables**, **points** and **division**, into the box designated **Block 1 of 1**. It is standard practice to leave the **Method** setting at the default **Enter**, which ensures that all specified predictors are included simultaneously in the model. Finally, click **OK** to execute the analysis and generate the output tables.



## Interpreting Key Output Metrics

Upon clicking **OK**, the [SPSS](#) Viewer window will populate with several tables detailing the results of the [logistic regression](#). Critical interpretation begins by examining the Model Summary and the Variables in the Equation tables. These sections provide the statistical evidence needed to assess the model's overall fit and the individual contributions of the predictors.

**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	8.256 <sup>a</sup>	.540	.725

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

**Classification Table<sup>a</sup>**

Observed		Predicted		Percentage Correct
		0	1	
Step 1	draft	5	1	83.3
		1	7	87.5
Overall Percentage				85.7

a. The cut value is .500

**Variables in the Equation**

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	points	.277	.134	4.243	1	.039	1.319
	division	-1.082	1.843	.345	1	.557	.339
	Constant	-3.152	3.408	.855	1	.355	.043

a. Variable(s) entered on step 1: points, division.

The **Model Summary** table offers an assessment of how well the model explains the variability in the outcome. The most informative metric here is the **Nagelkerke R Square**, which serves as a pseudo R-squared value for logistic models. This value indicates the percentage of variation in the **response variable** that can be explained by the inclusion of the predictor variables. In this specific case, the value suggests that the variables **points** and **division** collectively explain **72.5%** of the variability in the outcome of whether a player is drafted. This is considered a substantial explanatory power for a social science model.

The **Variables in the Equation** table is arguably the most important output, as it provides the specific statistical contribution of each predictor. Key metrics derived from this table include:

**Wald:** This is the **Wald test statistic** used to evaluate the null hypothesis that the population coefficient (B) for a predictor is zero. A significant result suggests the predictor variable should be retained in the model.

**Sig:** This column presents the **p-value** corresponding to the Wald test statistic for each **predictor variable**. We observe that the p-value for **points** is .039, which is below the standard significance threshold of .05, indicating that points per game is a statistically significant predictor. Conversely,

the p-value for **division** is .557, suggesting that it is not significant in this model.

**Exp(B):** This crucial metric represents the [odds ratio](#) for each predictor. The odds ratio quantifies the multiplicative change in the odds of the outcome event (getting drafted) associated with a one-unit increase in the predictor variable.

Based on the [odds ratio](#) results, we can draw specific conclusions. The odds of a player from division 2 getting drafted are only 0.339 times the odds of a player from division 1 being drafted (holding points constant). Furthermore, each single-unit increase in points per game is associated with an increase of 1.319 in the odds of a player getting drafted. These ratios provide actionable insights into the relationship between performance and professional outcome.

## Formulating the Predictive Equation and Reporting Results

The coefficients (B values) provided in the **Variables in the Equation** table are utilized to construct the full predictive equation for the probability of the outcome. This equation leverages the coefficients found in the column labeled B, allowing us to estimate the likelihood (P) that any given player will be drafted based on their specific performance metrics. The formula for calculating this probability using the logit transformation is as follows:

$$P(\text{Drafted}) = e^{(B_0 + B_1X_1 + B_2X_2)} / (1 + e^{(B_0 + B_1X_1 + B_2X_2)})$$

Substituting the specific [logistic regression coefficients](#) from the [SPSS](#) output (where B0 is the constant, B1 is the coefficient for points, and B2 is the coefficient for division), the operational formula becomes:

$$\text{Probability} = e^{-3.152 + .277(\text{points}) - 1.082(\text{division})} / (1 + e^{-3.152 + .277(\text{points}) - 1.082(\text{division})})$$

To illustrate the practical application of this formula, consider a player who averages 20 points per game (points = 20) and plays in division 1 (division = 1). We calculate their probability of being drafted:

$$\text{Probability} = e^{-3.152 + .277(20) - 1.082(1)} / (1 + e^{-3.152 + .277(20) - 1.082(1)}) = .787.$$

Since this calculated probability (0.787) exceeds the typical classification threshold of 0.5, the model strongly predicts that this specific player would be drafted into the NBA. This demonstrates the model's utility in generating specific, quantifiable predictions.

Lastly, reporting the results of the [logistic regression](#) should follow a clear, concise format, typically summarizing the model fit, the statistical significance of the predictors, and the practical implications derived from the [odds ratio](#). A comprehensive report ensures that the findings are accurately communicated to a broader audience:

A Binary Logistic Regression was executed to determine the influence of points per game and collegiate division level on a basketball player's probability of being drafted professionally. The analysis utilized data from 14 college players.

The resulting model demonstrated strong explanatory power, accounting for 72.5% of the variation in the draft outcome (as measured by the [Nagelkerke R Square](#)) and correctly classifying 85.7% of all cases in the dataset.

Statistically significant effects were observed for points per game. The [odds ratio](#) analysis indicated that the odds of a player in division 2 being drafted were only 0.339 times the odds for a player in division 1 (holding points constant).

Furthermore, the data showed that each incremental unit increase in points per game correlated with a significant increase of 1.319 in the odds of a player being selected in the professional draft.