

How to Perform Logistic Regression in SPSS: A Step-by-Step Guide

Authored by
stats writer

March 15, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform Logistic Regression in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135864>

Understanding the Fundamentals of Logistic Regression Analysis

In the realm of predictive analytics, **logistic regression** stands as a cornerstone technique for modeling the relationship between a set of predictor variables and a categorical outcome. Unlike linear regression, which is designed to predict continuous numeric values, logistic regression is specifically engineered to estimate the probability of an event occurring within a **binary** or dichotomous framework. This makes it an indispensable tool for researchers and data scientists who need to understand the likelihood of specific results, such as whether a customer will churn, whether a medical treatment will be successful, or, in the context of sports analytics, whether an athlete will be drafted into a professional league.

The mathematical foundation of this method relies on the logit transformation, which maps a linear combination of **independent variables** to a probability value ranging strictly between 0 and 1. This is achieved through the use of a logistic function, which produces an S-shaped curve rather than a straight line. By utilizing this approach, researchers can identify which factors are statistically significant and how they influence the "odds" of the **dependent variable** taking on a value of 1 versus 0. This distinction is crucial because it allows for the analysis of non-linear relationships that are common in real-world social and biological phenomena.

When performing this analysis, it is essential to consider the assumptions that underpin the model's validity. Although logistic regression does not require the strict normality of residuals that linear models do, it still necessitates that the observations be independent and that there is little to no **multicollinearity** among the predictors. Furthermore, the relationship between the independent variables and the logit of the outcome should be linear. Understanding these theoretical nuances ensures that the results derived from the software are both robust and generalizable to the broader population from which the sample was drawn.

Applying these concepts within a professional environment requires a reliable software package that can handle complex computations while providing a user-friendly interface. **SPSS** (Statistical Package for the Social Sciences) is widely regarded as one of the premier tools for this purpose. It offers a comprehensive suite of diagnostic tools and output tables that simplify the interpretation of coefficients, significance levels, and model fit. By following a structured workflow in SPSS, users can transform raw data into actionable insights with high precision and clarity.

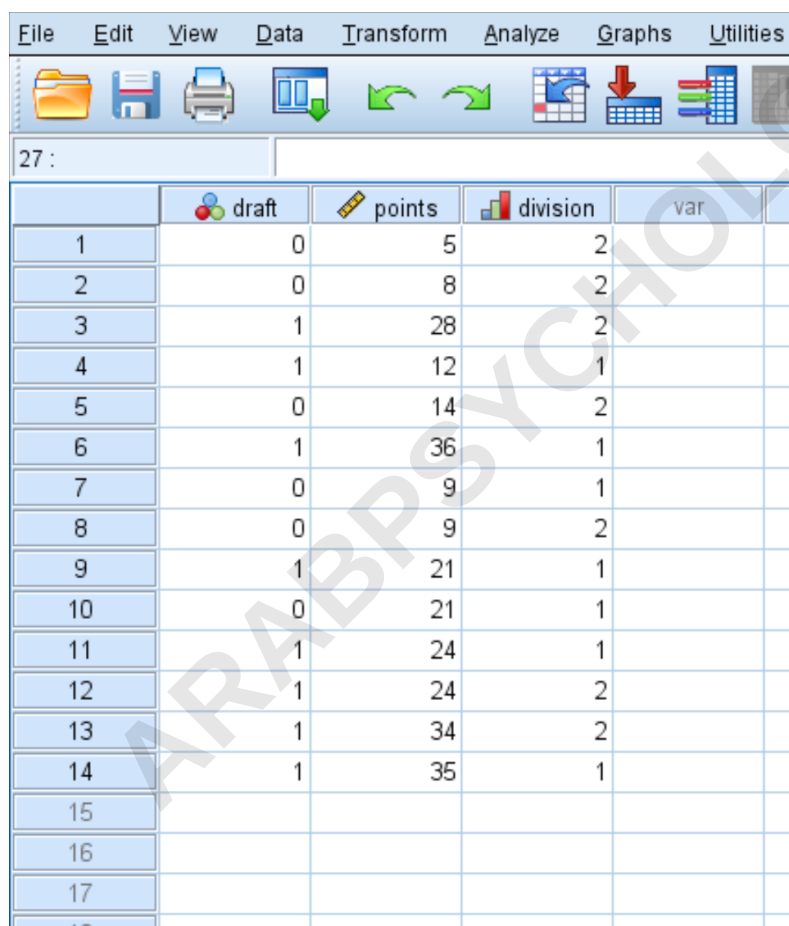
Data Preparation and Initial Input in SPSS

Before initiating the computational process, the first and perhaps most critical stage is the proper preparation and input of the dataset into the **SPSS** environment. Accuracy in data entry and variable definition is paramount, as any errors in the raw data will inevitably propagate through the analysis, leading to biased or incorrect conclusions. In our specific example, we are investigating the factors that influence whether college basketball players are drafted into the NBA. This requires

a dataset where each row represents an individual player and each column represents a specific attribute or outcome variable.

To begin, you must define your variables in the "Variable View" tab, ensuring that the outcome variable is coded appropriately. In this scenario, the variable "draft" serves as our binary outcome, where a value of 0 indicates the player was not drafted and a value of 1 indicates they were. Similarly, we must input our predictors: "points," representing the average points per game, and "division," indicating the level of competition. These **independent variables** provide the quantitative and qualitative basis upon which the model will build its predictions.

Once the variables are defined, the data points should be entered into the "Data View" grid. For this exercise, use the following visual guide to ensure your dataset matches the required structure for the regression analysis:



	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				

Effective data management also involves checking for missing values and ensuring that the measurement scales are correctly assigned. For instance, "points" should be treated as a scale variable, while "division" and "draft" are **categorical variables**. By meticulously organizing the data at this stage, you create a solid foundation for the subsequent statistical tests, allowing **logistic**

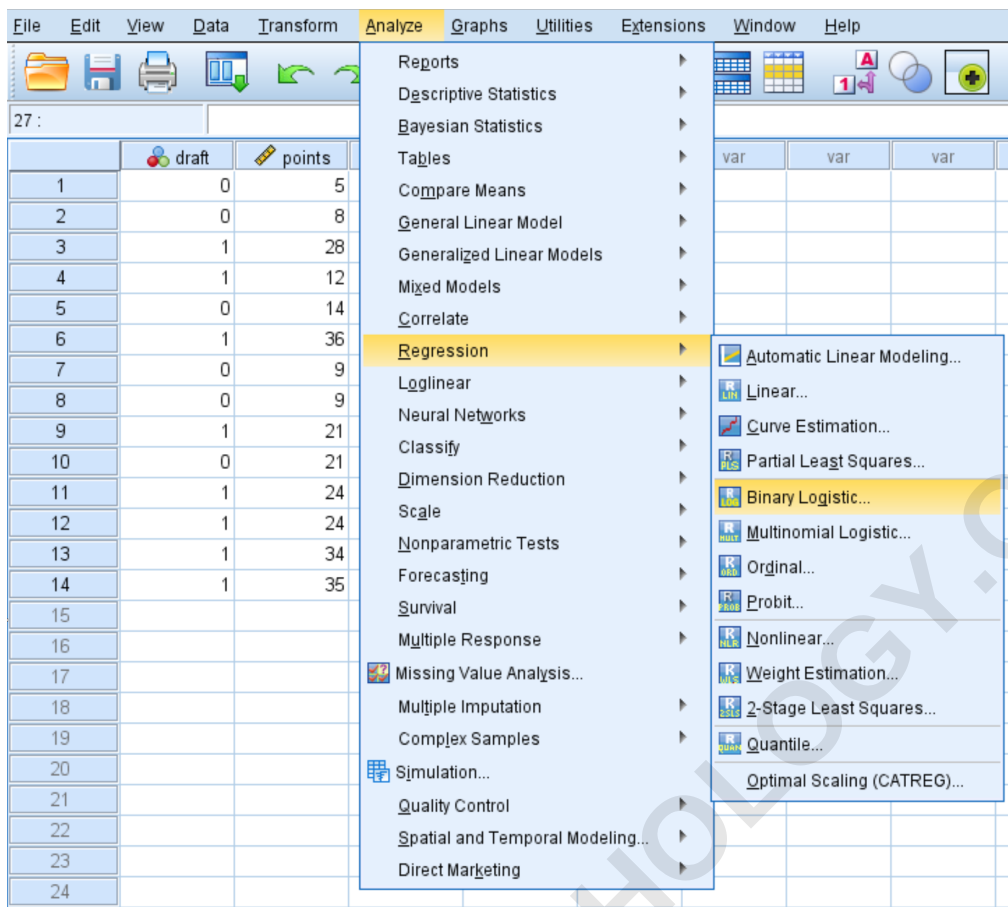
regression to yield the most accurate results possible regarding player draft probability.

Navigating the SPSS Menu for Binary Logistic Regression

With the dataset correctly formatted and loaded, the next phase involves navigating the **SPSS** menu system to execute the regression command. The software provides a dedicated module for **binary** outcomes, which automates the complex iterative estimations required to find the maximum likelihood estimates for the model coefficients. To access this feature, users should look to the top navigation bar, which houses the primary analytical functions of the program.

Specifically, you should click on the **Analyze** tab, which reveals a drop-down menu containing various statistical families. From there, hover over **Regression** and select **Binary Logistic** from the secondary list. This path is the standard gateway for performing **logistic regression** in SPSS, and it opens a specialized dialog box designed to help you specify the parameters of your model. The intuitive design of this interface allows for the rapid selection of variables without the need for manual coding or syntax writing.

The following image illustrates the precise menu path you need to follow within the software interface:



It is important to note that while the "Enter" method is the default and most commonly used approach, **SPSS** also offers stepwise methods for exploratory research. However, for most hypothesis-driven studies, the "Enter" method is preferred because it includes all specified predictors in the model simultaneously, allowing the researcher to evaluate the unique contribution of each variable while controlling for others. This step is fundamental in ensuring the integrity of the statistical comparison between variables like scoring average and division level.

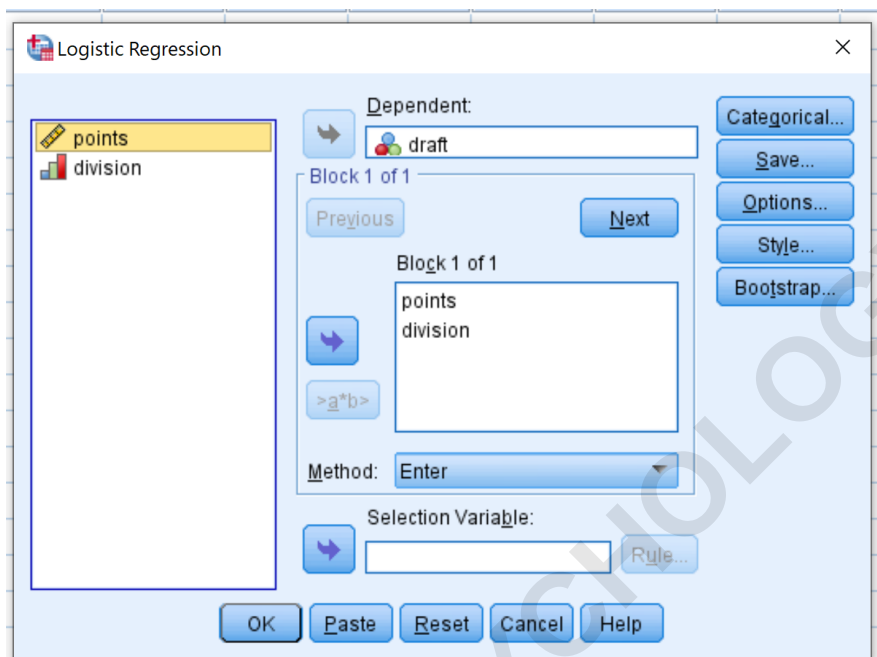
Configuring Model Variables and Options

Once the **Binary** Logistic Regression dialog box is active, you must assign the roles for each of your variables. This is the stage where you define the functional relationship you wish to test. The "Dependent" box is reserved for the outcome you are trying to predict--in this case, the "draft" variable. By placing it here, you are telling **SPSS** that this is the variable whose variance you intend to explain through the model. The software will then focus on predicting the probability that this variable equals 1.

Following this, you must identify your predictors and move them into the "Covariates" box. For our basketball analysis, you will drag "points" and "division" into this section. This tells the model to

treat these as the **independent variables** that influence the draft outcome. If you have categorical predictors with more than two levels, you might also need to click the "Categorical" button to define reference categories, though in our simplified two-division example, this may not be strictly necessary if the coding is already binary.

The visual representation of this configuration can be seen in the screenshot below, which serves as a template for your own analysis setup:



Before finalizing the setup by clicking **OK**, experienced researchers often explore the "Options" menu. This sub-menu allows you to request additional diagnostic statistics, such as the Hosmer-Lemeshow goodness-of-fit test, classification plots, and confidence intervals for the **odds ratio** (Exp(B)). These additional metrics provide a deeper level of validation, ensuring that the model not only calculates coefficients but also accurately reflects the underlying patterns in the data without overfitting or underfitting the sample.

Evaluating Model Summary and Goodness of Fit

After clicking the button to run the analysis, **SPSS** generates an output viewer containing several tables that are vital for interpretation. The first major area of focus should be the "Model Summary" table. This section provides a high-level overview of how well the regression model fits the observed data. Unlike linear models that use a standard R-squared, **logistic regression** utilizes "pseudo" R-squared values to estimate the amount of variance explained by the predictors.

One of the most frequently cited metrics in this table is the **Nagelkerke R Square**. This value is an

adjusted version of the Cox & Snell R Square and is designed to reach a maximum value of 1.0, making it easier to interpret as a percentage. In our basketball example, a Nagelkerke R Square of 0.725 indicates that approximately 72.5% of the variation in whether a player gets drafted can be explained by their points per game and the division they played in. This suggests a very strong model with high explanatory power.

Review the output table provided by the software to locate these specific metrics:

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	8.256 ^a	.540	.725

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

Classification Table^a

		Predicted		Percentage Correct
		0	1	
Step 1	draft 0	5	1	83.3
	1	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 ^a	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.

Beyond the R-squared values, the model summary also includes the -2 Log Likelihood, which is a measure of how poorly the model predicts the outcomes. A lower -2 Log Likelihood suggests a better fit. When comparing different models, researchers look for significant decreases in this value as new predictors are added. Understanding these fit indices is essential for determining if the overall model is statistically useful before diving into the significance of individual **independent variables**.

Interpreting the Variables in the Equation Table

The "Variables in the Equation" table is perhaps the most detailed and critical component of the

SPSS output. It breaks down the contribution of each predictor and provides the coefficients needed to construct the predictive formula. To interpret this table correctly, you must look at several key columns simultaneously: the B coefficients, the Wald statistic, the significance level, and the exponentiated B, commonly known as the **odds ratio**.

The **Wald test** is used to evaluate the null hypothesis that a particular coefficient is equal to zero. If the **p-value** (labeled "Sig." in SPSS) is less than the alpha level (typically 0.05), the variable is considered a significant predictor of the outcome. In our example, the p-value for "points" is .039, which is statistically significant. Conversely, "division" has a p-value of .557, suggesting that, when holding points constant, the division level does not have a statistically significant unique impact on the draft status in this specific, small dataset.

Key metrics found in this table include:

Wald: A statistical test that determines if the coefficient for a predictor is significantly different from zero.

Sig: The **p-value**; values below 0.05 indicate statistical significance for the predictor.

Exp(B): The **odds ratio**, which represents the change in odds resulting from a one-unit increase in the predictor.

B: The logistic regression coefficient, which is used in the probability calculation formula.

The **odds ratio** (Exp(B)) is particularly insightful for practical application. For the "points" variable, an Exp(B) of 1.319 implies that for every additional point per game a player averages, their odds of being drafted increase by approximately 31.9%. For the "division" variable, an Exp(B) of .339 suggests that players in division 2 have only about 33.9% of the odds of being drafted compared to those in division 1, though as noted, this specific finding did not reach statistical significance in our example.

Calculating and Applying Predicted Probabilities

The ultimate goal of **logistic regression** is often to make predictions about future or hypothetical cases. By using the B coefficients from the "Variables in the Equation" table, we can construct a mathematical formula to calculate the probability of the outcome for any given set of inputs. This process transforms the log-odds (logit) back into a standard probability percentage, which is much easier for stakeholders and decision-makers to understand.

Using the coefficients from our basketball model, the formula for probability is structured as follows: $\text{Probability} = e^{(\text{intercept} + B_1 \cdot x_1 + B_2 \cdot x_2)} / (1 + e^{(\text{intercept} + B_1 \cdot x_1 + B_2 \cdot x_2)})$. Plugging in our specific values (Intercept = -3.152, Points = .277, Division = -1.082), we can calculate the draft likelihood for a specific player profile. For a player averaging 20 points per game in Division 1, the calculation would look like this:

$$\text{Probability} = e^{(-3.152 + .277(20) - 1.082(1))} / (1 + e^{(-3.152 + .277(20) - 1.082(1))}) = 0.787$$

This result of **0.787** means there is a 78.7% probability that this specific player will be drafted. In **binary** classification, we typically use a cutoff of 0.5; since 0.787 is greater than 0.5, the model would formally predict that this player "will be drafted." This predictive capability is what makes **logistic regression** such a powerful tool for strategic planning and risk assessment across various industries.

Furthermore, **SPSS** can automate these calculations for every case in your dataset if you select the "Save" option before running the analysis and check the "Probabilities" box. This will add a new column to your data sheet containing the predicted probability for every single observation. This feature is invaluable for validating the model's accuracy against the actual observed outcomes and for identifying outliers or cases where the model's predictions were incorrect.

Synthesizing and Reporting Your Analysis Results

The final stage of the analytical workflow is the clear and concise reporting of the findings. In academic and professional reporting, it is not enough to simply list the numbers; you must interpret them in the context of the research question. A well-structured report should include the overall model fit, the significance of individual predictors, and the practical implications of the **odds ratios**. This ensures that the audience understands both the statistical validity and the real-world meaning of the data.

When summarizing your **logistic regression**, you should mention the total sample size (N=14 in our basketball example), the percentage of variance explained (Nagelkerke R Square = 72.5%), and the classification accuracy (85.7% of cases correctly classified). These figures provide an immediate sense of the model's reliability. Furthermore, you should highlight the specific impact of each significant variable, such as how each additional point per game increases the likelihood of being drafted.

A formal summary of our findings might be presented as follows:

Model Performance: The logistic regression model was statistically significant and explained 72.5% of the variation in draft outcomes, while correctly classifying 85.7% of the players in the sample.

Predictor Impact (Points): Points per game was a significant predictor ($p = .039$), with each unit increase in points per game associated with a 1.319 increase in the odds of being drafted.

Predictor Impact (Division): While division level showed that Division 2 players had lower odds (.339) than Division 1 players, this effect was not statistically significant in this sample ($p = .557$).

By presenting the results in this structured format, you provide a comprehensive view of the study

that is both scientifically rigorous and accessible. Whether you are writing a thesis, a peer-reviewed article, or a business intelligence report, following these guidelines for **logistic regression** in **SPSS** ensures that your data-driven insights are communicated with the highest degree of professionalism.

ARABPSYCHOLOGY.COM