

Learn How to Create and Interpret ROC Curves for Logistic Regression Analysis in SPSS

Authored by
Mohammed looti

November 8, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Create and Interpret ROC Curves for Logistic Regression Analysis in SPSS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12865>

[Logistic Regression](#) is a powerful statistical modeling technique fundamentally utilized when the dependent or response variable is [binary](#), meaning it has only two possible outcomes (e.g., success/failure, yes/no, drafted/not drafted). The primary goal of this model is to estimate the probability of the event occurring. However, simply fitting the model is not sufficient; we must rigorously assess its performance, particularly its ability to correctly classify outcomes. To evaluate how effectively a logistic regression model discriminates between the two categories within a dataset, we primarily rely on two critical, interconnected metrics that measure classification accuracy:

Sensitivity: Also known as the True Positive Rate (TPR), this is the probability that the model correctly predicts a positive outcome for an observation when the true outcome is, in fact, positive. High sensitivity is crucial when missing a positive case carries a high cost.

Specificity: Also known as the True Negative Rate (TNR), this is the probability that the model correctly predicts a negative outcome for an observation when the true outcome is negative. High specificity is essential when falsely identifying a negative case as positive is particularly detrimental.

These two measures, sensitivity and specificity, often exhibit a trade-off relationship, meaning that improving one frequently comes at the expense of the other. Finding the optimal balance between these metrics is vital for practical application of the model. The most effective and intuitive method for visualizing this fundamental trade-off across all possible classification thresholds is by constructing a [ROC curve](#) (Receiver Operating Characteristic curve). This plot is essential for understanding the overall discriminatory power of the predictive model.

This comprehensive tutorial will guide you through the process of generating and interpreting the results associated with a robust ROC curve using [SPSS Statistics](#), ensuring you can effectively evaluate your classification models.

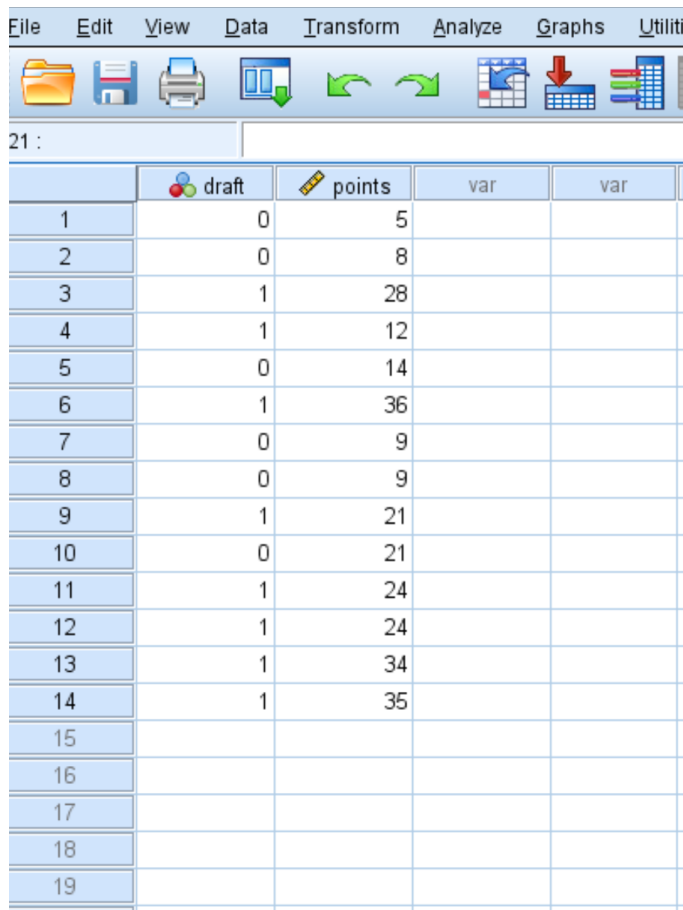
Example: Establishing the Foundation for ROC Analysis in SPSS

To illustrate the practical application of the ROC curve, let us consider a typical classification problem in sports analytics. Suppose we have gathered a dataset concerning a group of college basketball players. This dataset tracks whether or not each player was successfully drafted into the NBA, which serves as our binary response variable (coded as 0 for 'not drafted' and 1 for 'drafted'). Additionally, we have included a key potential predictor variable: their average points per game (PPG) recorded during their college career. We aim to determine how well PPG predicts the likelihood of being drafted and visualize this predictive power.

The structure of this dataset is crucial for the ROC analysis. The dependent variable, 'draft', must be clearly defined with its positive state identified, which in this case is the value '1'. The independent variable, 'points', is the continuous test variable against which the model's

discriminatory ability is assessed. Analyzing this relationship provides a clear, real-world context for applying the statistical concepts of sensitivity and specificity, allowing us to build a predictive model based on a single continuous measurement.

Below is a visual representation of the raw data as entered into the SPSS environment. Notice the clear binary distinction in the 'draft' variable and the corresponding numerical values in the 'points' variable, which will serve as the basis for our analysis:

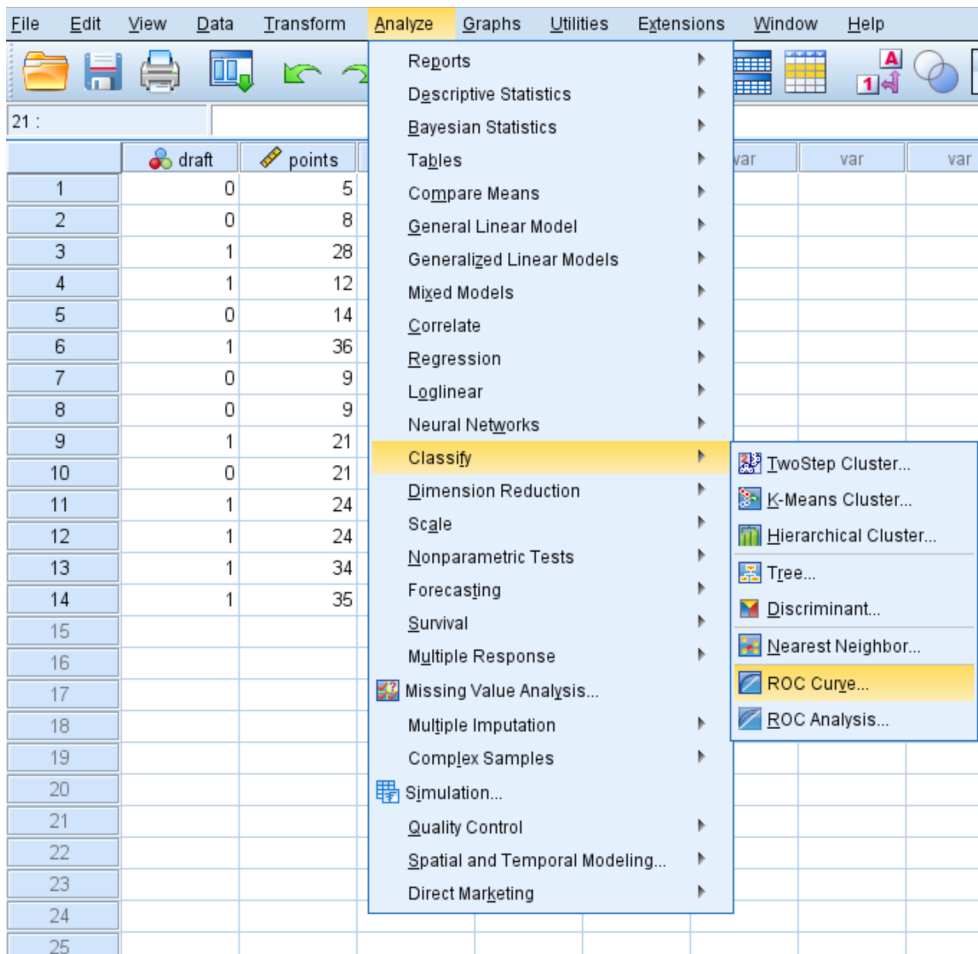


The screenshot shows the SPSS data editor window with a menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities) and a toolbar. The data grid contains 14 rows of data. The 'draft' column has values 0 or 1, and the 'points' column has values 5, 8, 28, 12, 14, 36, 9, 9, 21, 21, 24, 24, 34, 35. The remaining columns are labeled 'var'.

	draft	points	var	var
1	0	5		
2	0	8		
3	1	28		
4	1	12		
5	0	14		
6	1	36		
7	0	9		
8	0	9		
9	1	21		
10	0	21		
11	1	24		
12	1	24		
13	1	34		
14	1	35		
15				
16				
17				
18				
19				

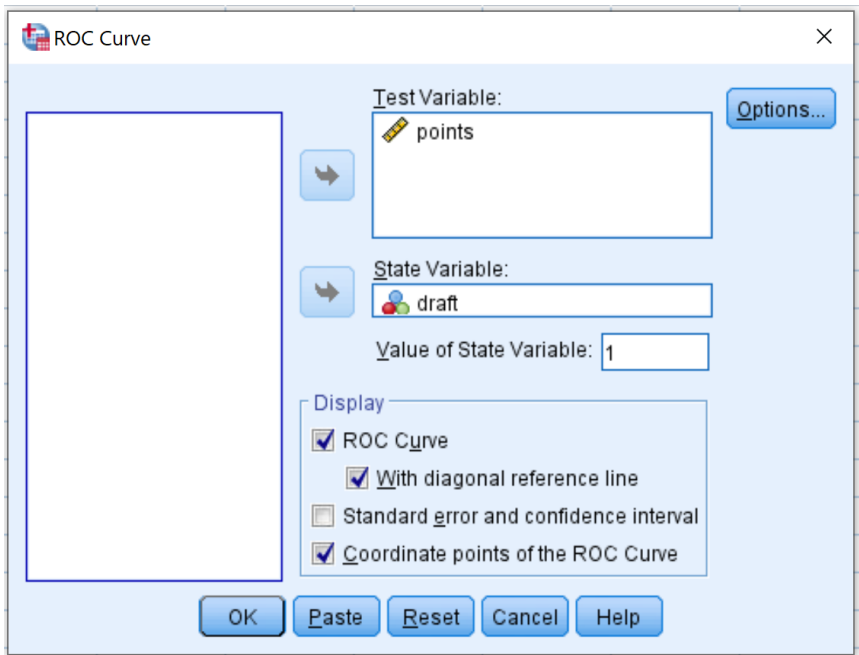
Generating the ROC Curve via the SPSS Menu

Once the data is correctly entered and verified within SPSS, generating the ROC curve is a straightforward process requiring just a few simple steps through the graphical user interface. This method is computationally efficient and produces all necessary output tables and visualizations for comprehensive evaluation. To begin the procedure, navigate to the main menu bar, click the **Analyze** tab, hover over **Classify**, and then select the **ROC Curve** option. This sequence initiates the specialized dialogue box tailored for this specific diagnostic analysis.



Within the newly opened dialogue box, you must correctly assign the variables to their designated roles. First, drag the dependent variable, **draft**, into the box labeled **State Variable**. Critically, you must then click the **Define Value of State Variable** button and explicitly set this value to **1**. This action tells SPSS that the value 1 represents the 'positive' outcome (getting drafted) that the model is attempting to predict. Next, drag the predictor variable, **points** (average points per game), into the box labeled **Test Variable**. This variable provides the continuous data used to calculate the classification probability.

To ensure the output is complete and easy to interpret, activate the following essential display options: check the box next to **With diagonal reference line**, which provides a benchmark for a random classifier, and also check **Coordinate points of the ROC Curve**. Requesting these coordinates is vital later on, as they allow us to examine the specific sensitivity and specificity values at various cut-off points. After confirming all variables and options are correctly set, proceed by clicking the **OK** button to execute the analysis and generate the ROC output.



Interpreting the Case Processing Summary and Visual Curve

The output generated by SPSS begins with the **Case Processing Summary**, which serves as a crucial starting point for verifying the data structure. This table meticulously displays the total counts of positive and negative cases that were utilized in the analysis. It confirms that the software correctly identified the outcomes based on the defined state variable (drafted = 1). In our basketball example, this summary reveals that 8 players were successfully drafted (positive results) and 6 players were not drafted (negative results). This step ensures that the sample size and outcome distributions are accurately accounted for before moving to the performance metrics.

ROC Curve

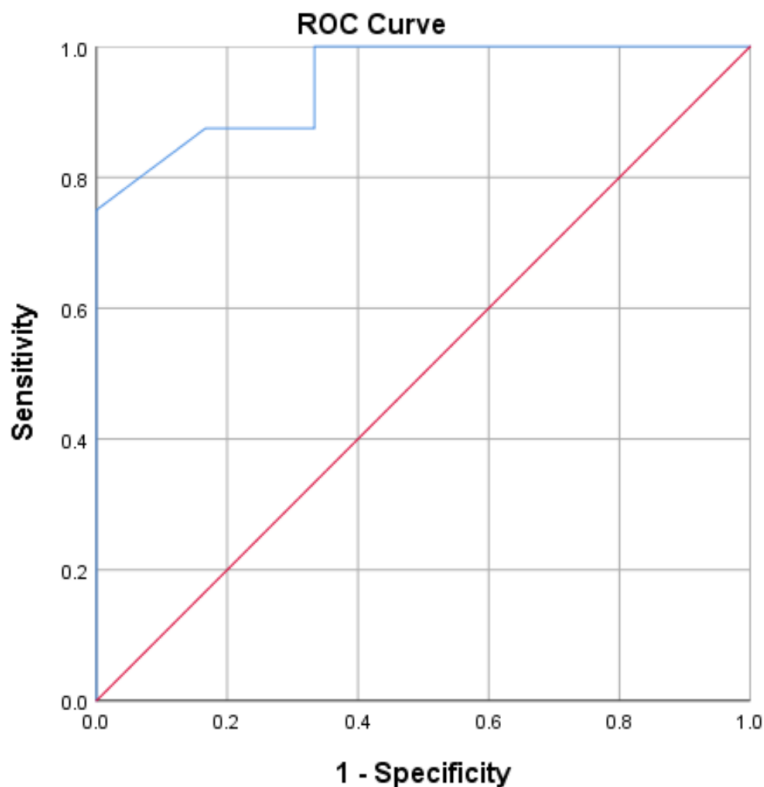
Case Processing Summary

draft	Valid N (listwise)
Positive ^a	8
Negative	6

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.

Following the summary, SPSS presents the graphical representation: the **ROC Curve** itself. This plot graphically displays the trade-off between sensitivity (True Positive Rate, plotted on the y-axis) and 1 - Specificity (False Positive Rate, plotted on the x-axis) across all possible classification thresholds. A perfectly performing model, one with maximum sensitivity and specificity, would yield a curve that passes through the top-left corner of the plot (coordinates 0, 1). Conversely, a useless model, one that performs no better than random chance, would result in a curve that lies directly on the 45-degree diagonal reference line.



Diagonal segments are produced by ties.

Observing the ROC curve (the blue line) generated for our specific model, we can immediately see that it closely "hugs" the top left boundary of the plot. This visual confirmation is highly encouraging, as it strongly suggests that the model developed using average points per game is an excellent discriminator. The vast distance between the blue curve and the diagonal reference line indicates that, across a wide range of thresholds, the model maintains a high true positive rate while simultaneously minimizing the false positive rate, thereby confirming its strong predictive performance in identifying which players will be drafted.

Analyzing the Area Under the Curve (AUC Metric)

While the visual assessment of the ROC curve provides an intuitive understanding of model performance, a quantitative metric is necessary for precise measurement and comparison. This metric is the [Area Under the Curve](#) (AUC), which provides a single, scalar value summarizing the model's ability to distinguish between the two classes (positive and negative outcomes). The AUC represents the probability that the model will rank a randomly chosen positive case higher than a randomly chosen negative case.

The AUC score ranges from 0 to 1. An AUC of 0.5 indicates that the model's performance is no better than random guessing (equivalent to the diagonal line). An AUC of 1.0 signifies a perfect

classifier that can distinguish between all positive and negative cases without error. Generally, an AUC value closer to 1.0 suggests a stronger, more reliable model. Standard interpretations often classify AUC scores between 0.7 and 0.8 as acceptable, 0.8 to 0.9 as excellent, and above 0.9 as outstanding. This metric is independent of the classification threshold chosen, making it an excellent measure of overall discriminatory power.

Area Under the Curve

Test Result Variable(s): points

Area

.948

The test result variable(s): points has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

For our specific logistic regression model predicting NBA drafting success, the AUC calculated by SPSS is reported as **.948**. This score is exceptionally high, placing the model firmly in the "outstanding" category. The numerical result reinforces the visual assessment derived from the curve itself, confirming that the average points per game is an extremely effective predictor variable. This high AUC value means that if we randomly select one drafted player and one non-drafted player, the model has a 94.8% chance of correctly assigning a higher probability score to the drafted player.

Deciphering the Coordinates of the Curve

The final and arguably most practical output table is the **Coordinates of the Curve**. This table details the specific sensitivity and 1 - specificity values corresponding to every unique cut-off point present in the dataset. While the AUC assesses the model globally, the coordinates table allows researchers to select the optimal threshold that best balances sensitivity and specificity for their particular application, based on the relative costs of false positives versus false negatives. Each row in this table represents a potential decision rule for classification.

Coordinates of the Curve

Test Result Variable(s): points

Positive if Greater Than or Equal To ^a	Sensitivity	1 - Specificity
4.00	1.000	1.000
6.50	1.000	.833
8.50	1.000	.667
10.50	1.000	.333
13.00	.875	.333
17.50	.875	.167
22.50	.750	.000
26.00	.500	.000
31.00	.375	.000
34.50	.250	.000
35.50	.125	.000
37.00	.000	.000

To understand the utility of these coordinates, let us examine a specific cut-off point, such as **8.50** points per game. If we choose this value as our classification threshold, our decision rule states that any player scoring less than 8.50 PPG is predicted to not be drafted, and any player scoring greater than 8.50 PPG is predicted to be drafted. The corresponding row in the coordinates table reveals the performance metrics for this specific decision rule.

Based on the coordinates table for a threshold of 8.50: our **sensitivity** is 100%. This means that the model successfully identified 100% of the truly positive cases (all drafted players were correctly predicted as drafted). The **1 - specificity** is 66.7%, which means our False Positive Rate is high. In this case, 66.7% of the players who were not drafted were incorrectly predicted as being drafted. Analyzing this table allows practitioners to carefully select a threshold--for instance, choosing one that maximizes sensitivity while keeping 1 - specificity below an acceptable level--to optimize the model's practical utility.