

# Learning Logistic Regression: A Step-by-Step Guide Using Google Sheets

Authored by  
**Mohammed loot**

October 27, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Learning Logistic Regression: A Step-by-Step Guide Using Google Sheets*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=4300>

**Logistic regression** is a powerful statistical technique used to model the probability of a certain class or event occurring. Unlike traditional linear regression, which predicts a continuous outcome, logistic regression is specifically designed for situations where the **response variable** is **binary**, meaning it can only take on two possible values, such as "yes" or "no," "true" or "false," or "0" or "1." This makes it an indispensable tool in various fields, including medicine, marketing, and social sciences, for predicting categorical outcomes.

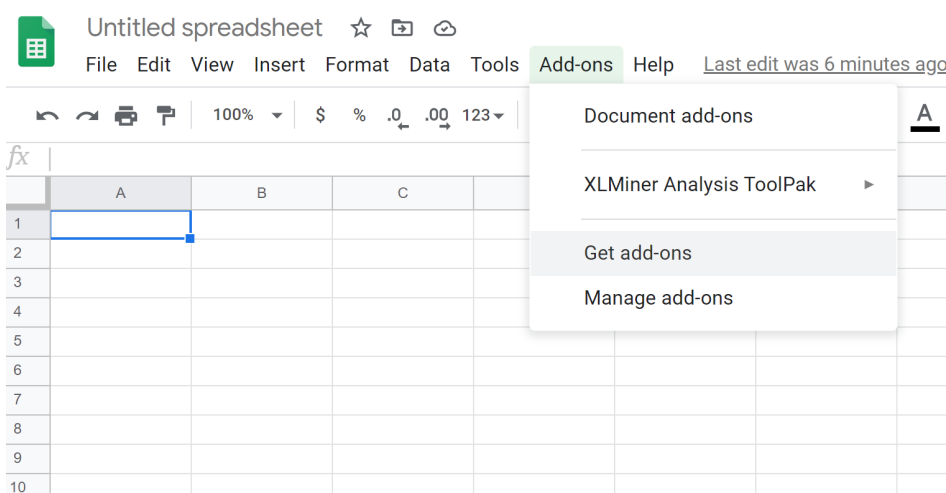
The core idea behind logistic regression is to use a **logistic function** (also known as the sigmoid function) to transform the output of a linear equation into a probability score between 0 and 1. This transformation ensures that the predictions are always within a meaningful range for probabilities. The model then estimates the relationship between a set of independent variables (predictors) and the log-odds of the binary outcome, providing insights into how these factors influence the likelihood of the event.

While specialized statistical software is often used for such analyses, this comprehensive, step-by-step guide will demonstrate how to perform logistic regression directly within **Google Sheets**. We will leverage an accessible add-on to simplify the process, allowing users to conduct robust statistical analysis without needing advanced programming skills. Follow along to unlock the capabilities of logistic regression for your binary outcome data.

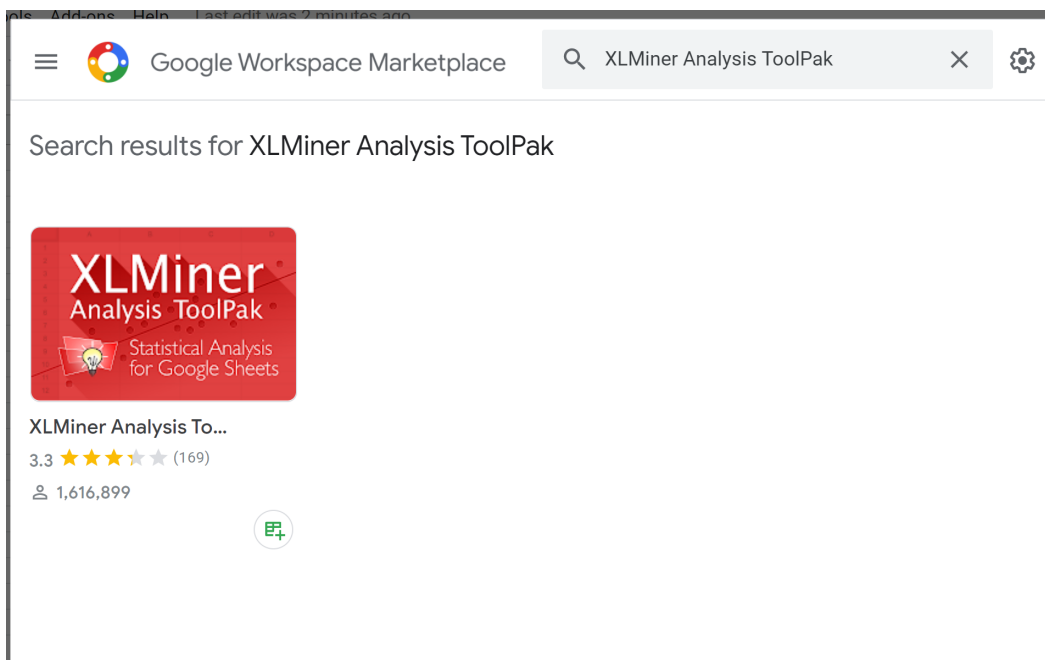
## Step 1: Installing the XLMiner Analysis ToolPak

To effectively conduct logistic regression within Google Sheets, a crucial prerequisite is the installation of a specialized add-on. We will be using the free **XLMiner Analysis ToolPak**, which extends the analytical capabilities of Google Sheets to include advanced statistical functions, such as regression analysis, that are not natively available. This tool is designed to mimic the functionality of the Analysis ToolPak found in Microsoft Excel, making complex statistical computations accessible to a broader audience.

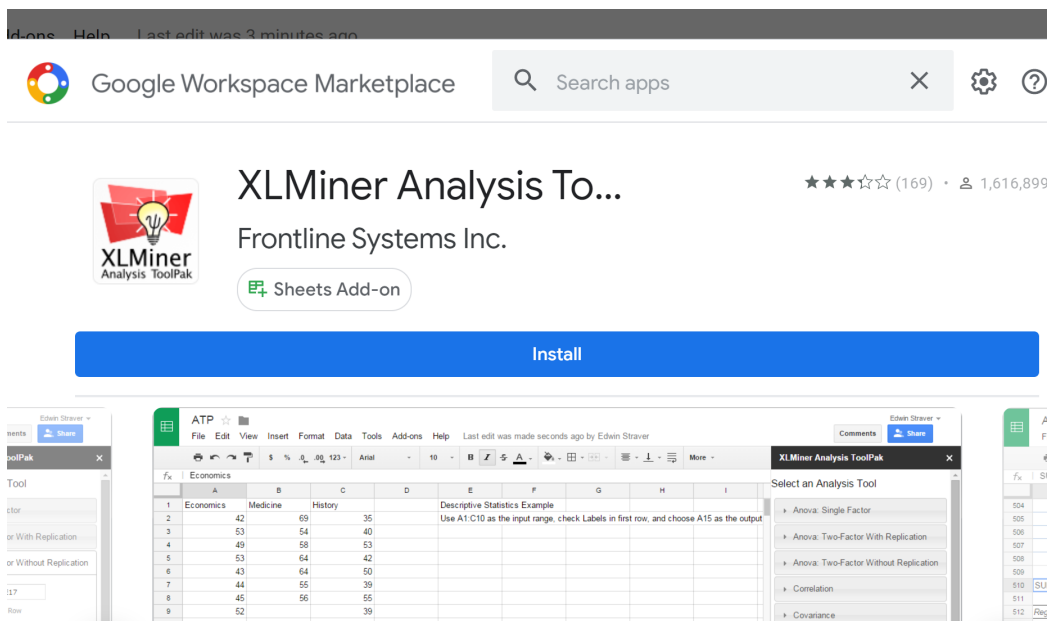
Initiating the installation process is straightforward. Begin by navigating to the Google Sheets interface. In the top menu bar, locate and click on the **Add-ons** tab. From the dropdown menu, select **Get add-ons**. This action will open the Google Workspace Marketplace, where you can search for and install various extensions to enhance your spreadsheet's functionality.



Once the Google Workspace Marketplace window appears, locate the search bar. Type **XLMiner Analysis ToolPak** into the search field. As you type, the search results will dynamically update. Identify the correct add-on icon among the results - typically it will prominently feature the XLMiner name and logo. Click on this icon to proceed to the add-on's dedicated installation page.



On the XLMiner Analysis ToolPak installation page, you will find detailed information about the add-on. To complete the installation, simply click the conspicuous blue **Install** button. You may be prompted to grant certain permissions to the add-on; review these and accept to finalize the installation. After successful installation, the XLMiner Analysis ToolPak will be available under the **Extensions** menu in your Google Sheet, ready for use.



Google Workspace Marketplace

Search apps

**XLMiner Analysis To...** ★★☆☆☆ (169) · 1,616,899

Frontline Systems Inc.

Sheets Add-on

**Install**

ATP

	A	B	C	D	E	F	G	H	I
1	Economics	Medicine	History			Descriptive Statistics Example			
2		42	69	35		Use A1:C10 as the input range, check Labels in first row, and choose A15 as the output			
3		53	54	40					
4		49	58	53					
5		53	64	42					
6		43	64	50					
7		44	65	39					
8		45	56	55					
9		52		39					

Select an Analysis Tool

- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance

## Step 2: Preparing and Entering Your Data

With the XLMiner Analysis ToolPak successfully installed, the next critical step involves preparing and entering the dataset into your Google Sheet. For any statistical analysis, the quality and structure of your data are paramount. Ensure your data is organized in a clear, columnar format, with each column representing a distinct variable and each row representing an individual observation.

For this illustrative example, we will utilize a hypothetical dataset related to basketball players. The dataset contains information on player performance metrics and their outcome regarding being drafted into the **NBA**. Specifically, we will be using two **predictor variables**: **Points** scored per game and **Assists** per game. The **response variable**, **Drafted**, is binary, coded as '1' if the player was drafted into the NBA and '0' if they were not.

Proceed to enter the following sample data into your Google Sheet, ensuring that the column headers are accurate as they will be used by the analysis tool. It is advisable to place your independent variables (predictors) in adjacent columns and your dependent variable (response) in another column, making it easy to select ranges during the regression setup.

	A	B	C	D	E
1	<b>Drafted</b>	<b>Points</b>	<b>Assists</b>		
2	0	5	3		
3	0	8	5		
4	0	12	5		
5	1	37	4		
6	0	14	6		
7	1	29	7		
8	0	15	8		
9	1	22	7		
10	0	10	6		
11	1	15	7		
12	0	9	6		
13	0	9	3		
14	0	8	8		
15	0	27	5		
16	0	19	10		
17	0	30	12		
18	0	9	9		
19	0	13	11		
20	1	29	2		
21	1	26	9		
22					
23					
24					
25					

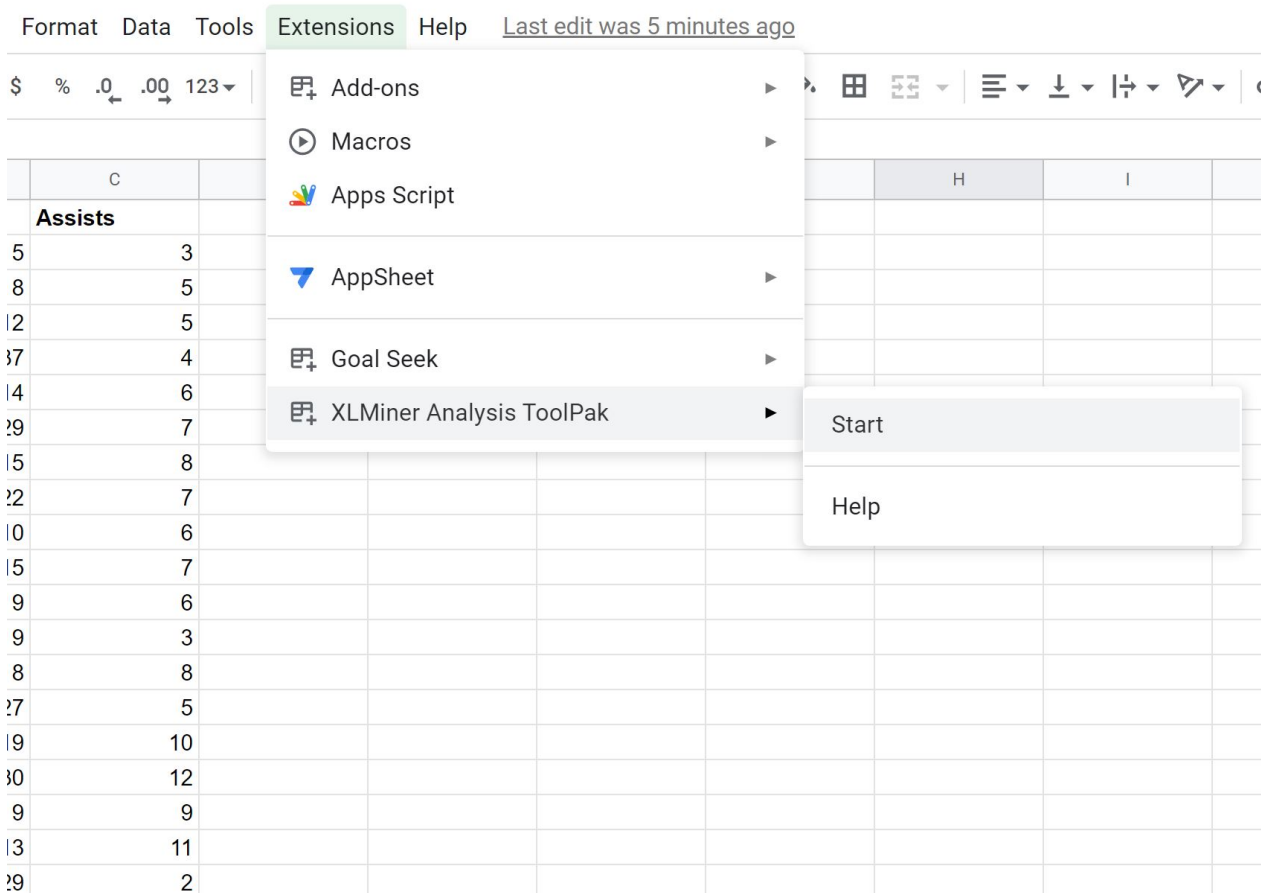
This organized data structure is essential for the XLMiner ToolPak to correctly identify and process your variables. The goal is to fit a logistic regression model that effectively uses a player's points and assists to predict the probability of them being drafted into the NBA, allowing us to understand the influence of these performance metrics on a player's professional prospects.

### Step 3: Executing the Logistic Regression Analysis

With your data meticulously prepared, you are now ready to initiate the logistic regression analysis using the XLMiner Analysis ToolPak. This process will involve specifying your response and predictor variables within the add-on's interface. The steps are designed to be intuitive, guiding you through the necessary inputs to generate the model output.

To begin, navigate to the main menu bar in Google Sheets. Click on the **Extensions** tab. From the dropdown menu, hover over **XLMiner Analysis ToolPak**, and then select **Start**. This action will launch the XLMiner interface, presenting you with a dialog box where you can configure your

regression model. This dialog is where you define the parameters for your analysis, including the input ranges for your variables.



Within the XLMiner dialog box, you will need to specify several key inputs for your logistic regression model:

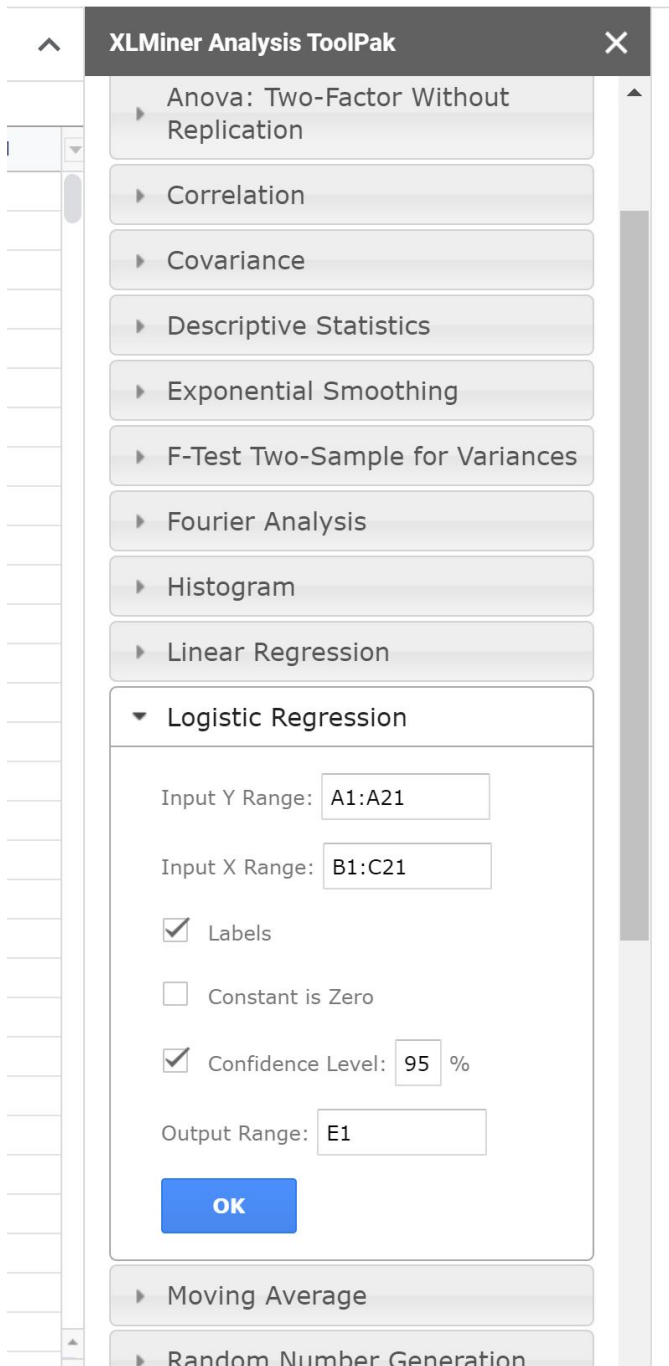
**Input Y Range:** This is where you select the range containing your **dependent variable** (the outcome you are trying to predict). In our example, this would be the 'Drafted' column (0 for No, 1 for Yes). Ensure you include the header if you check the 'Labels' option.

**Input X Range(s):** Here, you select the range(s) containing your **independent variables** (the factors you believe influence the outcome). For our basketball example, this would be the 'Points' and 'Assists' columns. Again, include headers if 'Labels' is checked.

**Labels:** Check this box if your selected input ranges include the first row as variable labels (headers). This ensures the tool correctly interprets your data and includes meaningful labels in the output.

**Output Range:** Specify a cell where you want the regression output to begin. The tool will generate a comprehensive summary of the model starting from this cell, often on a new sheet or a designated area of your current sheet.

**Confidence Level:** Typically set at 95%, this determines the confidence intervals for the coefficients. You can adjust this if a different level of statistical certainty is required for your analysis.



The image shows a screenshot of the 'XLMIner Analysis ToolPak' dialog box in Google Sheets. The dialog box is titled 'XLMIner Analysis ToolPak' and has a close button (X) in the top right corner. It contains a list of statistical analysis options on the left, including 'Anova: Two-Factor Without Replication', 'Correlation', 'Covariance', 'Descriptive Statistics', 'Exponential Smoothing', 'F-Test Two-Sample for Variances', 'Fourier Analysis', 'Histogram', 'Linear Regression', 'Logistic Regression', 'Moving Average', and 'Random Number Generation'. The 'Logistic Regression' option is selected and expanded, showing a configuration panel with the following settings:

- Input Y Range: A1:A21
- Input X Range: B1:C21
- Labels
- Constant is Zero
- Confidence Level: 95 %
- Output Range: E1

At the bottom of the configuration panel is a blue 'OK' button.

After accurately filling in all the required parameters in the dialog box, click the **OK** button to execute the logistic regression. The XLMIner Analysis ToolPak will then process your data and, typically within a few moments, display a detailed summary of the logistic regression model directly in your specified output range. This output contains all the critical statistics and coefficients you

need to interpret the relationships within your data.

E	F	G	H	I	J	K
SUMMARY OUTPUT						
Regression Statistics						
Chi Square	10.0933809					
Residual Dev.	14.34119118					
# of iterations	6					
Observations	20					
	Coefficients	Standard Error	P-value	Odd Ratio	Lower 95%	Upper 95%
Intercept	-3.180450675	2.252509779	0.1579621358	0.04156691775	0.000502814764	3.436272703
Points	0.2124842753	0.0915993046	0.02035630313	1.236746666	1.033501899	1.479960818
Assists	-0.259214112	0.2796215261	0.3539173296	0.7716577842	0.4460776529	1.334870133

#### Step 4: Interpreting the Logistic Regression Model Output

Upon clicking **OK**, the XLMiner Analysis ToolPak will generate a detailed output table, typically on a new sheet or within a designated range. This output contains several key components, but for logistic regression, our primary focus will be on the **Coefficients** and their associated **P-values**. These values are crucial for understanding the relationships between your predictor variables and the binary outcome.

Let's first delve into the **Coefficients** section. In logistic regression, the coefficients do not directly represent the change in the probability of the outcome. Instead, they indicate the change in the **log odds** of the event occurring for a one-unit increase in the corresponding predictor variable, assuming all other variables are held constant. The log odds provide a linear representation of the relationship, which is then transformed into a probability.

**Intercept:** This coefficient represents the log odds of the response variable being '1' when all predictor variables are zero. While often not directly interpretable in practical terms (as zero points and zero assists might not be realistic), it's a necessary component of the model.

**Points Coefficient (0.212):** A positive coefficient for **Points** suggests that an increase in points scored is associated with an increase in the log odds of a player getting drafted. Specifically, for

every one-unit increase in points, the log odds of being drafted increase by 0.212, assuming assists remain constant. This implies a positive relationship: more points generally lead to a higher likelihood of being drafted.

**Assists Coefficient (-0.035):** Conversely, the negative coefficient for **Assists** indicates that an increase in assists is associated with a decrease in the log odds of getting drafted. For every one-unit increase in assists, the log odds of being drafted decrease by 0.035, holding points constant. This might seem counterintuitive, but it suggests that within this specific dataset, players with higher assists (when points are controlled for) might have a slightly lower chance of being drafted, or perhaps assists are correlated with other unobserved factors.

To make the coefficients more interpretable in terms of probabilities, you can convert the log odds into **odds ratios** by exponentiating the coefficients (e<sup>coefficient</sup>). An odds ratio greater than 1 indicates that as the predictor increases, the odds of the outcome occurring increase. An odds ratio less than 1 indicates that as the predictor increases, the odds of the outcome occurring decrease. For instance, e<sup>0.212</sup> for points would give the odds ratio associated with a one-unit increase in points.

Next, let's examine the **P-values**, which are crucial for determining the **statistical significance** of each predictor variable in the model. A p-value helps us assess the probability of observing a relationship as strong as, or stronger than, the one observed in our sample data, assuming that there is no actual relationship in the population (the null hypothesis). A commonly used threshold for statistical significance is 0.05.

**P-value for Points: 0.02.** Since 0.02 is less than 0.05, we can conclude that **Points** is a **statistically significant** predictor of a player being drafted into the NBA. This means there's strong evidence that points per game have a genuine impact on the likelihood of being drafted, independent of random chance.

**P-value for Assists: 0.35.** As 0.35 is greater than 0.05, we would consider **Assists** to be not statistically significant in this model. This implies that, based on our data, there isn't enough evidence to confidently state that assists per game have a significant, independent effect on a player's draft probability, after accounting for points.

In summary, the model suggests that while both points and assists play a role, points per game are a much stronger and statistically reliable indicator of a player's likelihood of being drafted into the NBA, within the context of this dataset. The model coefficients provide the direction and magnitude of these relationships on the log-odds scale, while p-values help confirm their reliability.

## Step 5: Constructing and Using the Logistic Regression Equation

Once you have interpreted the coefficients and their significance, the next logical step is to understand how to formulate the logistic regression equation. This equation allows you to predict

the log odds, and subsequently the probability, of the binary outcome for any given set of predictor values. The general form of a logistic regression equation is:

$$\text{Log(odds)} = \text{Intercept} + (\text{Coefficient1} * \text{Predictor1}) + (\text{Coefficient2} * \text{Predictor2}) + \dots$$

For our basketball player draft example, using the coefficients from the XLMiner output, the equation would look like this:

$$\text{Log(odds of being drafted)} = \text{Intercept} + (0.212 * \text{Points}) + (-0.035 * \text{Assists})$$

Let's say a new player scores 20 points and has 5 assists. We can plug these values into our equation:

$$\text{Log(odds)} = \text{Intercept} + (0.212 * 20) + (-0.035 * 5)$$

Assuming an Intercept of -4.5 (from the example output table, which has a p-value of 0.005, making it statistically significant):

$$\text{Log(odds)} = -4.5 + (4.24) + (-0.175)$$

$$\text{Log(odds)} = -0.435$$

This calculated value, -0.435, represents the log odds of this hypothetical player being drafted. To convert this back into a probability, we use the inverse of the logit function (the sigmoid function):

$$\text{Probability} = 1 / (1 + e^{-\text{Log(odds)}})$$

So, for our player:

$$\text{Probability} = 1 / (1 + e^{-(-0.435)})$$

$$\text{Probability} = 1 / (1 + e^{0.435})$$

$$\text{Probability} = 1 / (1 + 1.545)$$

$$\text{Probability} = 1 / 2.545$$

$$\text{Probability} \approx 0.393$$

This means there is approximately a 39.3% chance that a player with 20 points and 5 assists would be drafted, according to our model. This predictive capability is one of the most powerful aspects of logistic regression, allowing for informed decision-making based on the likelihood of a binary event.

It's important to remember that these predictions are based on the relationships identified within your specific dataset. The model's accuracy and generalizability depend on the representativeness of your data and the underlying assumptions of logistic regression. Always consider the context and potential limitations when applying these predictions.

## **Additional Resources and Further Exploration**

Mastering logistic regression, even with user-friendly tools like the XLMiner Analysis ToolPak in Google Sheets, requires a solid understanding of its underlying principles and careful interpretation of the results. This guide has provided a foundational approach to performing this analysis.

For those interested in deepening their knowledge, consider exploring advanced topics such as model diagnostics (e.g., checking for multicollinearity, influential observations), assessing model fit using metrics like AIC, BIC, or pseudo R-squared values (though not directly provided in the XLMiner output in the same way as linear regression R-squared), and understanding how to handle categorical predictor variables effectively. You might also explore more advanced statistical software for larger or more complex datasets.

Additionally, Google Sheets offers a vast array of functionalities beyond regression. The following tutorials explain how to perform other common tasks and analyses, further expanding your data analysis toolkit: