

The 3 Types of Logistic Regression (Including Examples)

Authored by
Mohammed loot

November 3, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *The 3 Types of Logistic Regression (Including Examples)*.
PSYCHOLOGICAL STATISTICS. Retrieved from
<https://statistics.arabpsychology.com/?p=9066>

The technique known as [Logistic regression](#) is a cornerstone statistical and machine learning method widely employed across diverse fields, from epidemiology to financial modeling. Unlike its counterpart, [linear regression](#), this model is specifically engineered to handle situations where the outcome, or [response variable](#), is inherently **categorical** rather than continuous. Its primary function is to estimate the probability that a specific event will occur, based on the values of one or more independent variables.

A deep understanding of the nature of the categorical outcome is paramount, as the structure of the dependent variable dictates which specific variant of logistic regression must be implemented. While the core mathematical foundation remains consistent across all forms, the methodological application, interpretation of coefficients, and statistical assumptions vary significantly among the three principal types of the model.

Data scientists and statisticians classify logistic regression models into three primary categories, determined by the number of outcomes and the intrinsic properties of those response categories:

Binary logistic regression: Used when the outcome variable is restricted to two distinct, mutually exclusive options (e.g., presence/absence, pass/fail, true/false).

Multinomial logistic regression: Employed when the outcome variable contains three or more categories that possess absolutely **no natural order**, ranking, or hierarchy among them (e.g., colors, nationalities).

Ordinal logistic regression: Applied when the outcome variable consists of three or more categories that possess a natural, meaningful hierarchy or **ordering** (e.g., satisfaction ratings, severity levels).

The following visual aid effectively summarizes the structural differences that distinguish these powerful classification models:

Types of Logistic Regression Models

	Binomial Logistic Regression	Multinomial Logistic Regression	Ordinal Logistic Regression
Number of Categories for Response Variable	2	3 or more	3 or more
Does Order of Categories Matter?	No	No	Yes

This comprehensive guide will systematically explore each model type. We will provide detailed explanations of their applicability, discuss the specific structural assumptions that define them, and illustrate their utility through clear, real-world examples. Understanding these distinctions is crucial

for successfully tackling any classification challenge in data science.

Type #1: Binary Logistic Regression

[Binary logistic regression models](#) represent the most frequently encountered and fundamental application of the technique. These models are exclusively dedicated to situations where the [response variable](#) is dichotomous--meaning it can only assume one of two possible states, typically coded as 0 (absence) and 1 (presence, or the event of interest). The core objective is to model the probability of the outcome being '1' given a specific combination of input features.

To ensure that the output is always a valid probability, falling strictly between 0 and 1, this model utilizes the [sigmoid function](#) (also known as the logistic function). This function transforms the linear combination of the predictor variables into a probability score. Due to its simplicity, efficiency, and robust nature, binary logistic regression serves as the foundational technique for a vast array of [classification](#) tasks across industries, including clinical trials, manufacturing quality control, and detection of financial anomalies. Essentially, it provides a definitive answer to a yes/no question: Is the event likely to occur, or is it not?

Example 1: Predicting Professional Draft Status

Consider a sports data scientist attempting to identify which college basketball players possess the highest probability of successfully transitioning into the professional league. The scientist uses collegiate statistics--such as average points, rebounds, and assists--as independent variables. The critical [response variable](#) is whether a player is 'Drafted' (1) or 'Not Drafted' (0).

Since the potential outcomes are strictly limited to two distinct possibilities, the appropriate modeling technique is **binary logistic regression**. This allows the scientist to generate a predicted probability score for every player, quantifying their likelihood of selection in the professional draft based solely on their performance metrics.

Example 2: Automated Spam Detection Systems

A large telecommunications company needs to refine its email filtering system. The system's main task is to classify incoming messages into two mutually exclusive categories: 'Spam' or 'Not Spam' (often referred to as 'Ham'). They decide to leverage various predictors, such as the email's total word count, the frequency of specific keywords, and the sender's registered country of origin.

The clear division of the [response variable](#) into two states (spam vs. not spam) makes this scenario a textbook application of **binary logistic regression**. The resulting model efficiently estimates the probability that a new, unclassified email should be flagged as malicious or unsolicited content, thereby automating and improving user protection.

Type #2: Multinomial Logistic Regression

Multinomial logistic regression models are employed when the [categorical](#) response variable has three or more levels. The fundamental defining characteristic of this model is that the categories are strictly **nominal**; this means there is no inherent hierarchy, ranking, or logical sequence among the various potential outcomes. If the category labels were arbitrarily rearranged, the structural meaning of the model would remain unchanged.

This complex method operates by simultaneously estimating a set of binary logistic regressions, typically by comparing every category against a predetermined baseline or reference category. If the model involves K categories, it will estimate K-1 separate log-odds equations. The final output provides a set of probabilities for each individual, indicating the likelihood of them belonging to each specific category given the values of the predictor variables.

The critical distinction between the multinomial model and the ordinal model is the assumption of non-order. If an analyst mistakenly applies a multinomial model to truly ordered data, they sacrifice statistical power and efficiency. Conversely, attempting to impose an artificial order onto truly nominal data can yield statistically valid but practically nonsensical results. Multinomial models are indispensable tools for analyzing choices, preferences, and affiliations where all options are distinct and carry equal weight in terms of magnitude or value.

Example 1: Analyzing Political Preference

Imagine a political scientist seeking to understand the socio-economic factors influencing voter choice in an election featuring four major candidates (A, B, C, and D). The scientist uses independent variables like annual income, years of education, and demographic region to predict which candidate an individual is most likely to support.

Because the response variable has four options, and there is no inherent ranking (Candidate A is not statistically or logically "superior" to Candidate B), the appropriate methodology is **multinomial logistic regression**. This model accurately calculates the likelihood of a voter selecting each of the four candidates based on their data, producing four separate probabilities that collectively sum to one for every voter analyzed.

Example 2: Determining Consumer Brand Choice

A market research analyst investigates consumer purchasing behavior for common household appliances, where the options are three major brands: Zolt, Apex, or Vista. The analyst uses demographic predictors, such as household size and advertising exposure per week, to inform the prediction of brand preference.

Since the three brand preferences are distinct outcomes with no quantitative ranking (e.g., preferring Apex is not intrinsically "more" or "less" than preferring Zolt), the analyst must utilize **multinomial logistic regression**. This model efficiently handles the multiple, unordered choices, allowing the analyst to segment the market based on predicted interest in each brand.

Type #3: Ordinal Logistic Regression

Ordinal logistic regression models, often referred to as cumulative link models or proportional odds models, are required when the [categorical](#) response variable includes three or more levels that possess a clear, inherent, and meaningful sequence or order. Crucially, while the categories are ordered, the distances between them are not assumed to be equal, distinguishing them from truly continuous variables.

The primary statistical benefit of the ordinal model, compared to the multinomial approach which discards ordering information, is that it leverages this structural hierarchy, resulting in a more concise and statistically powerful model. This type of regression typically models the cumulative probabilities of observing an outcome up to or including a certain category level, rather than modeling the probability of falling into a single category.

The defining structural requirement of this method is the [Proportional Odds Assumption](#) (sometimes called the parallel slopes assumption). This assumption stipulates that the relationship between the predictor variables and the odds of the outcome being above a certain threshold is consistent across all potential category splits. Common applications are frequently found in contexts that rely on rating scales, such as customer satisfaction surveys (e.g., poor, fair, good, excellent), academic grading, or clinical severity scores.

Example 1: Predicting University Admission Caliber

An academic admissions advisor is developing a predictive model to forecast the caliber of university a prospective student will be admitted to. The universities are categorized into four ordered tiers: "Tier 4 (Lowest)," "Tier 3," "Tier 2," and "Tier 1 (Highest)." The advisor uses the student's academic metrics, such as GPA, standardized test scores, and extracurricular achievements, as [predictor variables](#).

Since the four classifications clearly possess a natural progression or ranking--Tier 4 is definitively lower than Tier 3, and Tier 1 is the highest--this scenario necessitates the use of an **ordinal logistic regression model**. This model efficiently uses the ordering information, recognizing that the factors influencing admission to a Tier 2 school versus a Tier 3 school are structurally related to the factors influencing admission to a Tier 1 school versus a Tier 2 school.

Example 2: Forecasting Customer Satisfaction Scores

A major service provider analyzes customer feedback data to predict future satisfaction levels based on service wait times and the complexity of the issue resolved. The satisfaction scale is recorded using five ordered levels: "Very Dissatisfied," "Dissatisfied," "Neutral," "Satisfied," and "Very Satisfied."

Given that the outcomes are multiple (five levels) and inherently ordered (a 'Satisfied' rating is superior to a 'Neutral' rating), the analyst must employ an **ordinal logistic regression model**. This model will assess the cumulative probability of a customer achieving a rating *at or above* a certain threshold (e.g., the probability of being 'Neutral' or better), leveraging the sequential nature of the score to achieve optimal predictive accuracy.

Choosing the Appropriate Logistic Regression Model

Selecting the correct type of logistic regression is far more than a simple statistical formality; it critically influences the validity of the model's underlying assumptions and the practical interpretability of its coefficients. Misapplying a model--for instance, using a binary model on a multi-class problem, or ignoring the crucial order in an ordinal setting--will inevitably lead to statistically inefficient results and potentially misleading business conclusions.

The fundamental decision hinges entirely on the structure and properties of the [response variable](#). If the response is limited exclusively to two outcomes, the choice is unequivocally **Binary logistic regression**. If, however, there are three or more outcomes, the analyst must then rigorously assess whether a natural, meaningful sequence or hierarchy exists among those options.

If the outcomes are nominal labels without any intrinsic rank (e.g., types of cars, political parties, or countries of origin), **Multinomial logistic regression** is the necessary tool. Conversely, if the outcomes represent a level, intensity, or severity that can be ranked (e.g., low/medium/high, academic grades, or Likert scales), **Ordinal logistic regression** is the appropriate, efficiency-maximizing choice, provided that the critical proportional odds assumption is met by the data.

Additional Resources for Deepening Your Knowledge

To further explore the mathematical underpinnings, practical implementation, and diagnostic checks required for these techniques, we highly recommend consulting specialized tutorials and official documentation on statistical modeling: