Glossary

alternatives. The set of alternatives are the possible choices a decision maker can pick or rank.

alternative-specific variable. When a variable varies across alternatives, it is called alternative specific. An alternative-specific variable may vary across alternatives only or across both alternatives and cases.

alternatives variable. A numeric or string variable that identifies the alternatives. Some models require an alternatives variable and some do not.

balanced. When choice sets are the same for every case, we say that they are balanced.

case. This is a Stata term for the set of Stata observations representing a single decision. A case contains one observation for each of the possible alternatives that the decision maker could have chosen or ranked.

case ID variable. A variable that identifies the cases. For independent cross-sectional data, this variable identifies the decision makers.

case-specific variable. When a variable is constant within a case, it is called case specific.

choice set. The set of alternatives a decision maker could have chosen or ranked. The choice sets can vary across cases.

discrete choice. When each decision maker picks a single alternative from his or her set of possible alternatives, it is called a discrete choice.

independence of irrelevant alternatives (IIA). The IIA property is true when adding another alternative to the set of alternatives does not change the relative probabilities of choosing alternatives from the initial set of alternatives.

observation. For choice models in Stata, there is a difference between Stata observations and statistical observations. We call a statistical observation a case. When we refer to an observation, we mean a Stata observation—one row in the dataset.

panel data. When decision makers make multiple choices at different time points, the data are panel data. A panel variable identifies decision makers, and a time variable identifies the time points.

rank-ordered alternatives. When each decision maker ranks his or her possible alternatives, we say we have rank-ordered alternatives.

unbalanced. When choice sets are not the same for every case, we say that they are unbalanced.

utility. Choice models are typically formulated using a latent continuous variable, called the utility, for each alternative. The largest value of the utility for each case represents the alternative chosen for discrete choices. For rank-ordered alternatives, the ranking of the values of the utilities gives the rank ordering of the choices.

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on citing Stata documentation.