What are the steps for conducting Probit Regression using Mplus for data analysis?

Authored by

stats writer

June 29, 2024

Probit Regression is a statistical method used for analyzing data with binary or categorical outcomes. Mplus is a popular software program that can be used to conduct Probit Regression analysis.

The following are the steps for conducting Probit Regression using Mplus:

1. Specify the model: The first step is to clearly define the research question and the variables that will be included in the analysis. This will help in selecting the appropriate Probit Regression model in Mplus.

2. Prepare the data: The data must be organized in a specific format that is compatible with Mplus. This includes creating a data file with the dependent and independent variables, as well as any control variables.

3. Specify the Probit Regression model: In Mplus, the Probit Regression model can be specified using the "MODEL" command. This command allows the researcher to specify the dependent and independent variables, as well as the type of Probit model to be used.

4. Estimate the model: Once the Probit model is specified, it can be estimated using the "ANALYSIS" command in Mplus. This will generate output that includes the model fit statistics, parameter estimates, and significance levels.

5. Interpret the results: The output from Mplus can be used to interpret the results of the Probit Regression analysis. The parameter estimates indicate the relationship between the independent variables and the probability of the outcome variable.

6. Evaluate the model fit: It is important to assess the fit of the Probit Regression model to the data. Mplus provides various fit statistics, such as the chi-square test, to evaluate the model fit.

7. Conduct sensitivity analysis: Sensitivity analysis can be used to assess the robustness of the Probit Regression results. This involves testing the model with different specifications or excluding certain variables.

8. Draw conclusions: Based on the results and model fit, conclusions can be drawn about the relationship between the independent variables and the probability of the outcome variable.

In summary, conducting Probit Regression using Mplus involves specifying the model, preparing the data, estimating the model, interpreting the results, evaluating the model fit, conducting sensitivity analysis, and drawing conclusions. These steps can help researchers accurately analyze binary or categorical data using Probit Regression in Mplus.

Probit Regression | Mplus Data Analysis Examples

Note: This example was done using Mplus version 5.2. The syntax may not work, or may function differently, with other versions of Mplus.

Probit regression, also called a probit model, is used to model dichotomous or binary outcome variables. In the probit model, the inverse standard normal distribution of the probability is modeled as a linear combination of the predictors.

Please note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics and potential follow-up analyses.

Examples

Example 1: Suppose that we are interested in the factors that influence

whether a political candidate wins an election. The outcome (response) variable
is binary (0/1); win or lose. The predictor variables of interest are the
amount of money spent on the campaign, the amount of time spent campaigning
negatively and whether the candidate is an incumbent.

Example 2: A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect
admission into graduate school. The response variable, admit/don't admit, is a
binary variable.

Description of the data

For our data analysis below, we are going to expand on Example 2 about
getting into graduate school. We have generatedhypotheticaldata,
which can be obtained by clicking on https://stats.idre.ucla.edu/wp-content/uploads/2016/02/binary.dat. You can store this anywhere you like, but our

examples will
assume it has been stored in c:data. (Note that the names of
variables should NOT be included at the top of the data file. Instead, the
variables are named in the variable command.) You may want to do your
descriptive statistics in a general use statistics package, such as SAS, Stata
or SPSS, because the options for obtaining descriptive statistics are limited in
Mplus. Even if you chose to run descriptive statistics in another package, it is
a good idea to run a model with type=basic before you do anything else,
just to make sure the dataset is being read correctly.

Data:
File is C:datahttps://stats.idre.ucla.edu/wp-content/uploads/2016/02/binary.dat ;
Variable:
Names are admit gre gpa rank rank1 rank2 rank3 rank4;
Analysis:

Type = basic;

As we mentioned above, you will want to look at this carefully to be sure that
the dataset was read into Mplus correctly. You will want to make sure that
you have the correct number of observations, and that the variables all have
means that are close to those from the descriptive statistics generated in a
general purpose statistical package. If there are missing values for some or all
of the variables, the descriptive statistics generated by Mplus will not match
those from a general purpose statistical package exactly, because by default, Mplus versions
5.0 and later use maximum likelihood based procedures for handling missing
values.

<output omitted>

SUMMARY OF ANALYSIS

Number of groups 1

Number of observations 400

<output omitted>

SAMPLE STATISTICS

Means
ADMIT GRE GPA RANK RANK1

_____ _____ _____ _____ _____

1 0.318 587.700 3.390 2.485 0.152

Means
RANK2 RANK3 RANK4

_____ _____ _____

1 0.378 0.302 0.168

Analysis methods you might consider

Below is a list of some analysis methods you may have encountered.
Some of the methods listed are quite reasonable while others have either
fallen out of favor or have limitations.

Using the probit model

The Mplus input file for a probit regression model is

shown below. Because
the data file contains variables that are not used in the model, the usevariables subcommand
is used to list the variables that are used in the model (i.e., admit,
gre, gpa, rank1, rank2 and rank3). Note that because Mplus uses the names subcommand to determine the order of variables
in the data file, the number and order of variables in the names subcommand should not be changed unless the data file is also changed. The
categorical subcommand is used to identify binary and ordinal outcome
variables. Categorical predictor
variables should be included as a series of dummy variables (e.g., rank1,
rank2, and rank3). We do not need to specify that we wish to run a
probit model because probit models are the default for binary outcome variables. Finally,
under model we specify that the outcome (i.e., admit) should be regressed
on the predictor variables (i.e., gre, gpa, rank1, rank2

and rank3).

Data:
File is D:documentsdae_updatinghttps://stats.idre.ucla.edu/wp-content/uploads/2016/02/binary.dat ;
Variable:
names = admit gre gpa rank rank1 rank2 rank3 rank4;
categorical = admit;
usevariables = admit gre gpa rank1 rank2 rank3;
Model:
admit on gre gpa rank1 rank2 rank3;

SUMMARY OF ANALYSIS

Number of groups 1
Number of observations 400

Number of dependent variables 1
Number of independent variables 5
Number of continuous latent variables 0

Observed dependent variables

Binary and ordered categorical (ordinal)

ADMIT

Observed independent variables
GRE GPA RANK1 RANK2 RANK3

Estimator WLSMV
Maximum number of iterations 1000
Convergence criterion 0.500D-04
Maximum number of steepest descent iterations 20
Parameterization DELTA

Input data file(s)
C:datahttps://stats.idre.ucla.edu/wp-content/uploads/2016/02/binary.dat

Input data format FREE

SUMMARY OF CATEGORICAL DATA PROPORTIONS

ADMIT
Category 1 0.683
Category 2 0.317

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value 0.000*
Degrees of Freedom 0**
P-Value 0.0000

*** The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used**
for chi-square difference tests. MLM, MLR and WLSM chi-square difference
testing is described in the Mplus Technical Appendices at www.statmodel.com.
See chi-square difference testing in the index of the Mplus User's Guide.

**** The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to**
a formula given in the Mplus Technical Appendices at www.statmodel.com.
See degrees of freedom in the index of the Mplus User's Guide.

Chi-Square Test of Model Fit for the Baseline Model

Value 33.821

Degrees of Freedom 5

P-Value 0.0000

CFI/TLI

CFI 1.000

TLI 1.000

Number of Free Parameters 6

RMSEA (Root Mean Square Error Of Approximation)

Estimate 0.000

WRMR (Weighted Root Mean Square Residual)

Value 0.005

MODEL RESULTS

Two-Tailed

Estimate S.E. Est./S.E. P-Value

ADMIT ON

GRE 0.001 0.001 2.122 0.034

GPA 0.478 0.189 2.529 0.011

RANK1 0.936 0.248 3.781 0.000

RANK2 0.520 0.211 2.464 0.014

RANK3 0.124 0.224 0.553 0.580

Thresholds

ADMIT$1 3.315 0.670 4.950 0.000

R-SQUARE

Observed Residual

Variable Estimate Variance

ADMIT 0.165 1.000

We can also test that the coefficients for rank1, rank2 and rank3, are all
equal to zero using the model test command. This type of test could also be
described as an overall test for the effect of rank. There are
multiple ways to test this type of hypothesis, the model test command
requests a Wald test. The Mplus input file shown
below is similar to the first model, except that the coefficients for rank1, rank2 and rank3

are assigned the names r1, r2 and r3, respectively. In the
model test command,
these coefficient names (i.e., r1, r2 and r3) are used to test that each of the coefficients is equal to 0.

Data:
File is C:datahttps://stats.idre.ucla.edu/wp-content/uploads/2016/02/binary.dat ;
Variable:
names = admit gre gpa rank rank1 rank2 rank3 rank4;
categorical = admit;
usevariables = admit gre gpa rank1 rank2 rank3;
Model:
admit on gre gpa
rank1 (r1)
rank2 (r2)
rank3 (r3);
Model test:
r1 = 0;
r2 = 0;
r3 = 0;

The majority of the output from this model is the same as the first model, so we will only show the part of the output that is associated with the model test command.

Wald Test of Parameter Constraints

Value 21.132
Degrees of Freedom 3
P-Value 0.0001

The portion of the output associated with the model test command is labeled "Wald Test of Parameter Constraints" and appears under the heading TESTS OF MODEL FIT. The test statistic is 21.132, with three degrees of freedom (one for each of the parameters tested), with an associated p-value of 0.0001. This indicates that the overall effect of rank is statistically significant.

We can also use the model test command to make pairwise comparisons among the terms for rank. The Mplus input below tests the hypothesis that

the coefficient for
rank2 (i.e., rank=2) is equal to the coefficient for rank3 (i.e.,
rank=3).

Data:
File is C:datahttps://stats.idre.ucla.edu/wp-content/uploads/2016/02/binary.dat;
Variable:
names = admit gre gpa rank rank1 rank2 rank3 rank4;
categorical = admit;
usevariables = admit gre gpa rank1 rank2 rank3;
Model:
admit on gre gpa
rank1 (r1)
rank2 (r2)
rank3 (r3);
Model test:
r2 = r3;

Below is the output associated with the model test command (as before, most of the model output is omitted).

Wald Test of Parameter Constraints

Value 5.682

Degrees of Freedom 1

P-Value 0.0171

Things to consider

References

Hosmer, D. & Lemeshow, S. (2000). Applied Logistic Regression (Second Edition). New York: John Wiley & Sons, Inc.

Long, J. Scott (1997). Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications.