“What is Probit Regression and how is it used in SPSS Data Analysis?”

Authored by

stats writer

June 29, 2024

Probit Regression is a statistical method used to analyze the relationship between a binary response variable and one or more independent variables. It is commonly used in data analysis to model the probability of an event or outcome occurring, based on the values of the independent variables. In SPSS, Probit Regression is a tool that allows users to estimate the probability of a binary response variable using a probit link function, which transforms the linear combination of the independent variables into probabilities. This method is useful in understanding the factors that influence the occurrence of a specific event or outcome, and can provide insights for decision making in various fields such as economics, social sciences, and public health.

Probit Regression | SPSS Data Analysis Examples

**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 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 generated hypothetical data, which can be
obtained by clicking on binary.sav. You can store this anywhere you like, but our examples will

assume it has been stored in c:data. First, we read the data file into SPSS.

```
get file = "c:dataprobit.sav".
```

This data set has a binary response (outcome, dependent) variable called admit. There are three predictor variables: gre, gpa and rank. We will treat the variables gre and gpa as continuous. The variable rank is ordinal, it takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. We will treat rank as categorical. Lets start by looking at descriptive statistics.

```
descriptives /variables=gre gpa.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation

gre 400 220 800 587.70 115.517

gpa 400 2.26 4.00 3.3899 .38057

Valid N (listwise) 400

frequencies /variables = rank admit.

Statistics

rank admit

N Valid 400 400

Missing 0 0

Frequency Table

rank

Frequency Percent Valid Percent Cumulative Percent

Valid 1 61 15.3 15.3 15.3

2 151 37.8 37.8 53.0

3 121 30.3 30.3 83.3

4 67 16.8 16.8 100.0

Total 400 100.0 100.0

admit

Frequency Percent Valid Percent Cumulative Percent

Valid 0 273 68.3 68.3 68.3

1 127 31.8 31.8 100.0

Total 400 100.0 100.0

crosstabs /tables = admit by rank.

Case Processing Summary

Cases

Valid Missing Total

N Percent N Percent N Percent

admit * rank 400 100.0% 0 .0% 400 100.0%

admit * rank Crosstabulation

Count

rank Total

1 2 3 4

admit 0 28 97 93 55 273

1 33 54 28 12 127

Total 61 151 121 67 400

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.

Probit regression

Below we use the plum command with the subcommand /link=probit to run a probit regression model.
After the command name (plum), the outcome variable (admit) is followed with

by rank which indicates that
rank is
a categorical predictor, followed by with gre gpa, indicating that the predictors
gre and gpa should be treated as continuous.

```
plum admit BY rank WITH gre gpa
/link=probit
/print= parameter summary.
```

The output from the plum command is broken into several sections, each of which is discussed below

Case Processing Summary

N Marginal Percentage

admit 0 273 68.3%

1 127 31.8%

rank 1 61 15.3%

2 151 37.8%

3 121 30.3%

4 67 16.8%

Valid 400 100.0%

Missing 0

Total 400

Model Fitting Information

Model -2 Log Likelihood Chi-Square df Sig.

Intercept Only 493.620

Final 452.057 41.563 5 .000

Link function: Probit.

Pseudo R-Square

Cox and Snell .099

Nagelkerke .138

McFadden .083

Link function: Probit.

Parameter Estimates

Estimate Std. Error Wald df Sig. 95% Confidence Interval

Lower Bound Upper Bound

Threshold 3.323 .663 25.090 1 .000 2.023 4.623

Location gre .001 .001 4.478 1 .034 .000 .003

gpa .478 .197 5.869 1 .015 .091 .864

.936 .245 14.560 1 .000 .455 1.417

.520 .211 6.091 1 .014 .107 .934

.124 .224 .305 1 .581 -.315 .563

0a . . 0 . . .

Link function: Probit.

a. This parameter is set to zero because it is redundant.

We may also want to test the overall effect of rank, we can do this using the test subcommand. The test subcommand is followed by the

name of the variable we wish
to test (i.e., rank), and then one value for each level of that
variable (including the omitted category). The first line of the test subcommand
rank 1 0 0 0 indicates that we want to test that the coefficient for
rank=1 is 0. To perform a multiple degree of freedom test, we include
multiple lines in the test subcommand, all but the last line is separated by a
semicolon. The second and third rows indicate that we wish to test that the
coefficients for rank=2 and rank=3 are equal to 0. Note that there is no need to
include a row for the fourth category of rank.

plum admit by rank with gre gpa
/link=probit
/print= parameter summary
/test rank 1 0 0 0;
rank 0 1 0 0;
rank 0 0 1 0.

Because the models are the same, most of the output produced by the above
plum command is the same as before. The only difference is the additional output
produced by the test subcommand, only this portion of the output is
shown below.

Custom Hypothesis Tests 1

Contrast Coefficients

C1 C2 C3

Threshold 0 0 0

Location gre 0 0 0

gpa 0 0 0

1 0 0

0 1 0

0 0 1

0 0 0

Contrast Results

Contrasts Estimate Std. Error Test value Wald df Sig. 95% Confidence Interval

Lower Bound Upper Bound

C1 .936 .245 0 14.560 1 .000 .455 1.417

C2 .520 .211 0 6.091 1 .014 .107 .934

C3 .124 .224 0 .305 1 .581 -.315 .563

Link function: Probit.

Test Results

Wald df Sig.

21.361 3 .000

Link function: Probit.

The table labeled Parameter Estimates gives hypothesis tests for differences between each level of rank and the reference category. We can use the

test subcommand to test for differences between the other levels of rank. For example, we might want to test for a difference in coefficients for rank=2 and rank=3.

In the syntax below we have added a second test subcommand. This time, the values given are 0 1 -1 0 this indicates that we want to calculate the difference between the coefficients for rank=2 and rank=3 (i.e., rank=2 - rank=3).

```
plum admit by rank with gre gpa
/link=probit
/print= parameter summary
/test rank 1 0 0 0;
rank 0 1 0 0;
rank 0 0 1 0
/test rank 0 1 -1 0.
```

Again the output from the model, as well as the output associated with the first test subcommand are identical to those shown above, so they are omitted.

Custom Hypothesis Tests 2

Contrast Coefficients

C1

Threshold 0

Location gre 0

gpa 0

0

1

-1

0

Contrast Results

Contrasts Estimate Std. Error Test value Wald df Sig. 95% Confidence Interval

Lower Bound Upper Bound

C1 .397 .168 0 5.573 1 .018 .067 .726

Link function: Probit.

In the table labeled Contrast Results we see the difference in the coefficients (i.e., 0.397).
The
Wald test statistic of 5.573, with one degree of freedom, and associated p-value
of less than 0.02, indicates that
the difference between the coefficients for rank=2 and rank=3 is
statistically significant. Because only one estimate was specified in the test
subcommand, the multiple degree of freedom test (i.e. the Test Results table) is
not printed.

Things to consider

See also

References