What is the process of conducting a Logit Regression analysis using SPSS?

Authored by

stats writer

June 29, 2024

Logit Regression analysis is a statistical method used to model the relationship between a categorical dependent variable and one or more independent variables. SPSS (Statistical Package for the Social Sciences) is a software program commonly used for data analysis. The process of conducting a Logit Regression analysis using SPSS involves several steps. First, the researcher must import the data into the SPSS program. Then, the researcher must specify the dependent and independent variables for the analysis. Next, the researcher must select the appropriate Logit Regression model based on the type of data and research question. The program will then generate the regression results, including the coefficient estimates, significance levels, and goodness of fit measures. The researcher can also use the output to interpret the results and make conclusions about the relationship between the variables. Finally, the researcher may choose to visualize the results through graphs or charts. Overall, conducting a Logit Regression analysis using SPSS involves importing the data, specifying variables, running the analysis, and interpreting the results to gain insights into the relationship between variables.

Logit Regression | SPSS Data Analysis Examples

Version info: Code for this page was tested in SPSS 20.

Logistic regression, also called a logit model, is used to model dichotomous outcome variables. In the logit model the log odds of the outcome is modeled as a linear combination of the predictor variables.

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 or not 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 from our website by clicking on binary.sav.
You can store this anywhere you like, but the syntax below assumes it has been
stored in the directory c:data.
This dataset has a binary response (outcome, dependent) variable called admit,
which is equal to 1 if the individual was admitted to graduate school, and 0
otherwise. There are three
predictor variables: gre, gpa, and rank. We will treat the variables
gre and gpa as continuous.
The variable rank 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 start out by opening

the dataset and looking at some descriptive statistics.

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

descriptives /variables=gre gpa.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
gre	400	220.00	800.00	587.7000	115.51654
gpa	400	2.26	4.00	3.3899	.38057
Valid N (listwise)	400				

```
frequencies /variables = rank.
```

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	

```
crosstabs /tables = admit by rank.
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
admit * rank	400	100.0%	0	0.0%	400	100.0%

admit * rank Crosstabulation

Count

		rank				
		1	2	3	4	Total
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.

Logistic regression

Below we use the logistic regression command to run a model predicting the outcome variable
admit, using gre, gpa, and rank. The categorical option specifies that rank is a categorical rather than continuous variable. The
output is shown in sections, each of which is discussed below.

logistic regression admit with gre gpa rank

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	400	100.0
	Missing Cases	0	.0
	Total	400	100.0
Unselected Cases		0	.0
Total		400	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

/categorical = rank.

The first table above shows a breakdown of the number of cases used and not used in the analysis. The second table above gives the coding for the outcome variable, admit.

Categorical Variables Codings[a]

			Parameter coding		
		Frequency	(1)	(2)	(3)
rank	1	61	1.000	.000	.000
	2	151	.000	1.000	.000
	3	121	.000	.000	1.000
	4	67	.000	.000	.000

a. This coding results in indicator coefficients.

The table above shows how the values of the

categorical variable

rank were handled, there are terms (essentially dummy variables) in the model for

rank=1,

rank=2, and rank=3; rank=4 is the omitted category.

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted		
			admit		Percentage Correct
			0	1	
Step 0	admit	0	273	0	100.0
		1	127	0	.0
	Overall Percentage				68.3

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.765	.107	50.764	1	.000	.465

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	gre	13.606	1	.000
		gpa	12.704	1	.000
		rank	25.242	3	.000
		rank(1)	16.590	1	.000
		rank(2)	1.801	1	.180
		rank(3)	5.934	1	.015
	Overall Statistics		40.160	5	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	41.459	5	.000
	Block	41.459	5	.000
	Model	41.459	5	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	458.517[a]	.098	.138

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			admit		Percentage Correct
	Observed		0	1	
Step 1	admit	0	254	19	93.0
		1	97	30	23.6
	Overall Percentage				71.0

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	gre	.002	.001	4.284	1	.038	1.002
	gpa	.804	.332	5.872	1	.015	2.235
	rank			20.895	3	.000	
	rank(1)	1.551	.418	13.787	1	.000	4.718
	rank(2)	.876	.367	5.706	1	.017	2.401
	rank(3)	.211	.393	.289	1	.591	1.235
	Constant	-5.541	1.138	23.709	1	.000	.004

a. Variable(s) entered on step 1: gre, gpa, rank.

Things to consider

See also

References