What is Multinomial Logistic Regression and how is it used in SPSS for data analysis?

Authored by

stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is Multinomial Logistic Regression and how is it used in SPSS for data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=157646

Multinomial Logistic Regression is a statistical technique used to predict the probability of an outcome with multiple categories. It is commonly used in data analysis to model relationships between a set of independent variables and a categorical dependent variable with more than two categories. In SPSS, it is used to analyze data with multiple response options, where the dependent variable is categorical and the independent variables are either continuous or categorical. This method allows for the inclusion of multiple predictor variables and can handle data with more than two response options, making it a powerful tool for analyzing complex data sets. By estimating the probability of each category, this technique can help identify the factors that influence the outcome and make predictions for new data. Overall, multinomial logistic regression is a valuable tool for understanding and analyzing relationships between variables in SPSS.

Multinomial Logistic Regression | SPSS Data Analysis Examples

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

Multinomial logistic regression is used to model nominal outcome variables, in which the log odds of the outcomes are 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 of multinomial logistic regression

Example 1. People's occupational choices might be influenced by their parents' occupations and their own education level. We can study the relationship of one's occupation choice with education level and father's occupation. The occupational choices will be the outcome variable which consists of categories of occupations.

Example 2. A biologist may be interested in food choices that alligators make. Adult alligators might have difference preference than young ones. The outcome variable here will be the types of food, and the predictor variables might be the length of the alligators and other environmental variables.

Example 3. Entering high school students make

program choices among general program, vocational program and academic program. Their choice might be modeled using their writing score and their social economic status.

Description of the data

For our data analysis example, we will expand the third example using the hsbdemo data set. You can download the data here.

The data set contains variables on 200 students. The outcome variable is prog, program type. The predictor variables are social economic status, ses, a three-level categorical variable and writing score, write, a continuous variable. Let's start with getting some descriptive statistics of the variables of interest.

get data "D:datahsbdemo.sav".

crosstabs

/tables=prog by ses

/statistics=chisq

/cells=count.

type of program * ses Crosstabulation

Count

		ses			Total
		low	middle	high	
type of program	general	16	20	9	45
	academic	19	44	42	105
	vocation	12	31	7	50
Total		47	95	58	200

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.604[a]	4	.002
Likelihood Ratio	16.783	4	.002
Linear-by-Linear Association	.060	1	.807
N of Valid Cases	200		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 10.58.

sort cases by prog.

split file by prog.

descriptives var = write

/statistics = mean stddev.

Descriptive Statistics

type of program		N	Mean	Std. Deviation
general	writing score	45	51.33	9.398
	Valid N (listwise)	45		
academic	writing score	105	56.26	7.943
	Valid N (listwise)	105		
vocation	writing score	50	46.76	9.319
	Valid N (listwise)	50		

split file off.

Analysis methods you might consider

Using the multinomial logit model

Below we use the nomreg command to estimate a multinomial logistic regression model. We specify the baseline comparison group to be the academic group using (base=2).

nomreg prog (base = 2) by ses with write /print = lrt cps mfi parameter summary.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	254.986			
Final	206.756	48.230	6	.000

Pseudo R-Square

Cox and Snell	.214
Nagelkerke	.246
McFadden	.118

Parameter Estimates

type of program[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)	
								Lower Bound	Upper Bound
general	Intercept	1.689	1.227	1.896	1	.169			
	write	-.058	.021	7.320	1	.007	.944	.905	.984
	[ses=1]	1.163	.514	5.114	1	.024	3.199	1.168	8.764
	[ses=2]	.630	.465	1.833	1	.176	1.877	.754	4.669
	[ses=3]	0[b]	.	.	0	.	.	.	.
vocation	Intercept	4.236	1.205	12.361	1	.000			
	write	-.114	.022	26.139	1	.000	.893	.855	.932
	[ses=1]	.983	.596	2.722	1	.099	2.672	.831	8.585
	[ses=2]	1.274	.511	6.214	1	.013	3.575	1.313	9.736
	[ses=3]	0[b]	.	.	0	.	.	.	.

a. The reference category is: academic.

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

$$\ln\left(\frac{P(prog=general)}{P(prog=academic)}\right) = b_{10} + b_{11}(ses=1) + b_{12}(ses=2) + b_{13}write$$

$$\ln\left(\frac{P(prog=vocation)}{P(prog=academic)}\right) = b_{20} + b_{21}(ses=1) + b_{22}(ses=2) + b_{23}write$$

where (b)'s are the regression coefficients.

The ratio of the probability of choosing one outcome category over the probability of choosing the baseline category is often referred to as relative risk (and it is also sometimes referred to as odds as we have just used to described the regression parameters above). Thus, exponentiating the linear equations above yields relative risks. Regression coefficients represent the change in log relative risk (log odds) per unit change in the predictor. Exponentiating regression coefficients will therefore yield relative risk ratios. SPSS includes relative risk ratios in the output, under the column "Exp(B)".

Tests for the overall effect of ses and write are outputted by the nomreg command. Below we see that the effects are

statistically significant.

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	206.756[a]	.000	0	.
write	238.203	31.447	2	.000
ses	217.815	11.058	4	.026

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

You can also use predicted probabilities to help you understand the model. You can calculate predicted probabilities using the SPSS matrix command. Below we calculate the predicted probability of choosing each program type at each level of ses, holding write at its means.

Matrix.

*** intercept1 intercept2 pared public gpa.**

```
* these coefficients are taken from the output.
compute b_gen = {1.689354 ; -0.057928 ; 1.162832 ; 0.629541}.
compute b_voc = {4.235530 ; -0.113603 ; 0.982670 ; 1.274063}.
* overall design matrix including means of public and gpa.
compute x = {{1 ; 1; 1}, make(3, 1, 52.775), {1, 0; 0, 1; 0, 0}}.
compute lp_gen = exp(x * b_gen).
compute lp_voc = exp(x * b_voc).
compute lp_aca = {1; 1; 1}.
compute p_gen = lp_gen/(lp_aca + lp_gen + lp_voc).
compute p_voc = lp_voc/(lp_aca + lp_gen + lp_voc).
compute p_aca = lp_aca/(lp_aca + lp_gen + lp_voc).
compute p = {p_gen, p_aca, p_voc}.
print p /title 'Predicted Probabilities for Outcomes 1 2 3 for ses 1 2 3 at mean of write'.
End Matrix.
```

Run MATRIX procedure:

Predicted Probabilities for Outcomes 1 2 3 for ses 1 2 3 at mean of write

.3581989665 .4396824687 .2021185647

.2283388262 .4777491509 .2939120229

.1784967500 .7009009604 .1206022896

------ END MATRIX -----

Column 1 contains the predicted probabilities for prog = general, where ses equals 1, 2 and 3 on each successive row. Columns 2 and 3 are the same for prog = academic and prog = vocational, respectively. We can also calculate predicted probabilities as we vary write from 30 to 70, when ses = 1.

Matrix.
*** intercept1 intercept2 pared public gpa.**
*** these coefficients are taken from the output.**
compute b_gen = {1.689354 ; -0.057928 ; 1.162832 ; 0.629541}.
compute b_voc = {4.235530 ; -0.113603 ; 0.982670 ; 1.274063}.
*** overall design matrix including means of public and gpa.**
compute x = {make(5,1,1), {30; 40; 50; 60; 70}, make(5,1,1), make(5,1,0)}.

compute lp_gen = exp(x * b_gen).
compute lp_voc = exp(x * b_voc).
compute lp_aca = {1; 1; 1; 1; 1}.
compute p_gen = lp_gen/(lp_aca + lp_gen + lp_voc).
compute p_voc = lp_voc/(lp_aca + lp_gen + lp_voc).
compute p_aca = lp_aca/(lp_aca + lp_gen + lp_voc).
compute p = {p_gen, p_aca, p_voc}.
print p /title 'Predicted Probabilities for Outcomes 1 2 3 for write 30 40 50 60 70 at ses=1'.
End Matrix.

Run MATRIX procedure:

Predicted Probabilities for Outcomes 1 2 3 for write 30 40 50 60 70 at ses=1
.2999966732 .0984378501 .6015654767
.3656613530 .2141424912 .4201961559
.3698577661 .3865775582 .2435646757
.3083735022 .5752505689 .1163759289
.2199925775 .7324300249 .0475773976

------ END MATRIX -----

Column 1 contains the predicted probabilities for prog = general, where write equals 30, 40, 50, 60 and 70 for

rows 1 through 5, respectively. Columns 2 and 3 are the same for prog = academic and prog = vocational, respectively.

Things to consider

See also

References