

What is the SAS annotated output for Multinomial Logistic Regression?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the SAS annotated output for Multinomial Logistic Regression?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160045>

The SAS annotated output for Multinomial Logistic Regression is a comprehensive summary of the statistical analysis results generated by the SAS software for a given dataset. It includes information such as the model parameters, standard errors, significance levels, and odds ratios, which provide insights into the relationship between the predictor variables and the outcome variable in a multinomial logistic regression model. The output also includes diagnostic measures and goodness-of-fit tests to assess the overall performance of the model. This annotated output serves as a valuable tool for researchers and analysts to interpret and communicate the results of a multinomial logistic regression analysis.

Multinomial Logistic Regression | SAS Annotated Output

This page shows an example of a multinomial logistic regression analysis with footnotes explaining the output. The dataset, mlogit, was collected on 200 high school students and are scores on various tests, including a video game and a puzzle. The outcome measure in this analysis is the preferred flavor of ice cream - vanilla, chocolate or strawberry- from which we are going to see what relationships exists with video game scores (video), puzzle scores (puzzle) and gender (female). Our response variable, ice_cream, is going to be treated as categorical under the assumption that the

levels of ice_cream

have no natural ordering, and we are going to allow SAS to choose the

referent group. In our example, this will be strawberry.

By default, SAS sorts

the outcome variable alphabetically or numerically and selects the last group to

be the referent group. The variable ice_cream is a numeric variable in

SAS, so we will add value labels using proc format.

```
data mlogit;
```

```
set "C:mlogit";
```

```
run;
```

```
proc format;
```

```
value ice_cream_l
```

```
1="chocolate"
```

```
2="vanilla"
```

```
3="strawberry";
```

```
run;
```

Before running the multinomial logistic regression, obtaining a frequency of

the ice cream flavors in the data can inform the selection of a reference group.

```
proc freq data = mlogit;  
format ice_cream ice_cream_l.;  
table ice_cream;  
run;
```

The FREQ Procedure

favorite flavor of ice cream

Cumulative Cumulative

ICE_CREAM	Frequency	Percent	Frequency	Percent
-----------	-----------	---------	-----------	---------

chocolate	47	23.50	47	23.50
-----------	----	-------	----	-------

vanilla	95	47.50	142	71.00
---------	----	-------	-----	-------

strawberry	58	29.00	200	100.00
------------	----	-------	-----	--------

We can use proc logistic for this model and indicate that the link

function is a generalized logit. This model allows for more than two categories

in the modeled variable and will compare each category to a reference category.

If we do not specify a reference category, the last

ordered category (in this case, ice_cream = 3) will be considered as the reference.

```
proc logistic data = mlogit;  
model ice_cream = video puzzle female / link = glogit;  
run;
```

Note that we could also use proc catmod for the multinomial logistic regression.

proc catmod is designed for categorical modeling and multinomial logistic

regression is an example of such a model. The options we would use within proc

catmod would specify that our model is a multinomial logistic regression. On

the direct statement, we can list the continuous predictor variables. On the

response statement, we would specify that the response functions are generalized logits. Finally, on

the model

statement, we would indicate our outcome variable ice_cream and the predictor

variables to be included in the model. See the proc catmod code below.

This yields an equivalent model to the proc logistic code above.

```
proc catmod data = mlogit;  
direct video puzzle female;  
response logits;  
model ice_cream = video puzzle female;  
run;
```

The output annotated on this page will be from the proc logistic commands.

The proc logistic code above generates the following output:

The LOGISTIC Procedure

Model Information

Data Set WORK.MLOGIT

Response Variable ICE_CREAM favorite flavor of ice cream

Number of Response Levels 3

Model generalized logit

Optimization Technique Fisher's scoring

Number of Observations Read 200

Number of Observations Used 200

Response Profile

Ordered Total

Value ICE_CREAM Frequency

1 1 47

2 2 95

3 3 58

Logits modeled use ICE_CREAM=3 as the reference category.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Intercept

Intercept and

Criterion Only Covariates

AIC 425.165 404.070

SC 431.762 430.456

-2 Log L 421.165 388.070

Testing Global Null Hypothesis: BETA=0

Test Chi-Square DF Pr > ChiSq

Likelihood Ratio 33.0954 6 ChiSq

VIDEO 2 3.4297 0.1800

PUZZLE 2 11.8188 0.0027

FEMALE 2 4.8352 0.0891

Analysis of Maximum Likelihood Estimates

Standard Wald

Parameter ICE_CREAM DF Estimate Error Chi-Square

Pr > ChiSq

Intercept 1 1 5.9691 1.4375 17.2425

Model Information

Data Summary

Data Set WORK.MLOGIT

Response Variablea ICE_CREAM favorite flavor of ice cream

Number of Response Levelsb 3

Model generalized logit

Optimization Technique Fisher's scoring

Number of Observations Readc 200

Number of Observations Usedc 200

a. ResponseVariable - This is the response variable in the model. For this example, the response variable is ice_cream.

b. Number of Response Levels - This indicates how many levels exist within the response variable. It also indicates how many models are fitted in the multinomial regression. In our dataset, there are three possible values for

ice_cream (chocolate, vanilla and strawberry), so there are three levels to our response variable. In a multinomial regression, one level of the response variable is treated as the referent group, and then a model is fit for each of the remaining levels compared to the referent group. Since we have three levels, one will be the referent level (strawberry) and we will fit

two models: 1)

chocolate relative to strawberry and 2) vanilla relative to strawberry.

c. Number of Observations Read/Used - The first is the number of observations in the model dataset. The second is the number of observations in the dataset with valid data in all of the variables needed for the specified model. In this example, our dataset does not contain any missing values, so the number of observations used in our model is equal to the number of observations read in from our dataset.

Response Profiles

Ordered Total

Value ICE_CREAM Frequency

1 1 47

2 2 95

3 3 58

Logits modeled use ICE_CREAM=3 as the reference

category.

d. Response Profiles - This outlines the order in which the values of our outcome variable ice_cream are considered. By default in SAS, the last value is the referent group in the multinomial logistic regression model. In this case, the last value corresponds to ice_cream = 3, which is strawberry. Additionally, the numbers assigned to the other values of the outcome variable are useful in interpreting other portions of the multinomial regression output.

Model Fit Statistics and Overall Tests of Effects

Intercept

Intercept and

Criteria Only Covariates

AIC 425.165 404.070

SC 431.762 430.456

-2 Log L 421.165 388.070

Testing Global Null Hypothesis: BETA=0**Test Chi-Square DF Pr > ChiSq****Likelihood Ratio 33.0954 6 <.0001****Score 30.5499 6 <.0001****Wald 26.8597 6 0.0002****Type 3 Analysis of Effects****Wald****Effect DF Chi-Square Pr > ChiSq****VIDEO 2 3.4297 0.1800****PUZZLE 2 11.8188 0.0027****FEMALE 2 4.8352 0.0891**

e. Criterion - These are various measurements used to assess the model

fit. The first two, Akaike Information Criterion (AIC) and Schwarz

Criterion (SC) are deviants of negative two times the Log-Likelihood (-2

Log L). AIC and SC penalize the Log-Likelihood by the number

of predictors in the model.

AIC - This is the Akaike Information Criterion. It is

calculated

as $AIC = -2 \text{ Log } L + 2((k-1) + s)$, where k is the number of levels of the dependent variable and s is the number of predictors in the model. AIC is used for the comparison of models from different samples or nonnested models. Ultimately, the model with the smallest AIC is considered the best.

SC - This is the Schwarz Criterion. It is defined as $-2 \text{ Log } L + ((k-1) + s) \cdot \log(\sum f_i)$, where f_i 's are the frequency values of the i th observation, and k and s were defined previously. Like AIC, SC penalizes for the number of predictors in the model and the smallest SC is most desirable.

-2 Log L - This is negative two times the log likelihood. The

-2 Log L is used in hypothesis tests for nested models.

f. Intercept Only - This column lists the values of the

specified fit

criteria from a model predicting the response variable without covariates (just an intercept).

g. Intercept and Covariates - This column lists the values of the specified fit criteria from a model predicting the response variable with the covariates indicated in the model statement.

h. Test - This indicates which Chi-Square test statistic is used to test the global null hypothesis that none of the predictors in either of the models have non-zero coefficients. The test statistics provided by SAS include the likelihood ratio, score, and Wald Chi-Square statistics.

i. Chi-Square - These are the values of the specified Chi-Square test statistics.

j. DF - These are the degrees of freedom for each of the

tests three

global tests. Since all three are testing the same hypothesis, the degrees of freedom is the same for all three. There are a total of six parameters (two models with three parameters each) compared to zero, so the degrees of freedom is 6.

k. $Pr > ChiSq$ - This is the p-value associated with the specified Chi-Square statistic. Here, the null hypothesis is that there is no relationship between the any of the predictor variable and the outcome, ice_cream (i.e., the estimates of the all of the predictors in both of the fitted models is zero). If the p-value is less than the specified alpha (usually .05 or .01), then this null hypothesis can be rejected. In this example, all three tests indicate that we can reject the null hypothesis.

l.

Effect - Here, we are interested in the effect of of each predictor on the outcome variable considering both of the fitted models at once.

m. DF -

The degrees of freedom for this analysis refers to the two fitted models, so $DF=2$ for all of the variables.

n. Wald Chi-Square -

This is the post-estimation test statistic of the parameter across both models.

o. $Pr > ChiSq$ - This is the p-value associated with the Wald Chi-Square statistic. Here, the null hypothesis is that there is no relationship between the predictor variable and the outcome, ice_cream (i.e., the estimates of the predictor in both of the fitted models are zero). If the p-value is less than the specified alpha (usually .05 or .01), then this null hypothesis can be

rejected.

Analysis of Maximum Likelihood Estimates

Analysis of Maximum Likelihood Estimates

Standard Wald

Parameter	ICE_CREAM	q	DF	Estimates	Error	Chi-Square	Pr > ChiSq
-----------	-----------	---	----	-----------	-------	------------	------------

Intercept	1	1	5.9691	1.4375	17.2425	<.0001
-----------	---	---	--------	--------	---------	--------

Intercept	2	1	4.0572	1.2229	11.0065	0.0009
-----------	---	---	--------	--------	---------	--------

VIDEO	1	1	-0.0465	0.0251	3.4296	0.0640
-------	---	---	---------	--------	--------	--------

VIDEO	2	1	-0.0229	0.0209	1.2060	0.2721
-------	---	---	---------	--------	--------	--------

PUZZLE	1	1	-0.0819	0.0238	11.8149	0.0006
--------	---	---	---------	--------	---------	--------

PUZZLE	2	1	-0.0430	0.0199	4.6746	0.0306
--------	---	---	---------	--------	--------	--------

FEMALE	1	1	0.8494	0.4482	3.5913	0.0581
--------	---	---	--------	--------	--------	--------

FEMALE	2	1	0.0328	0.3500	0.0088	0.9252
--------	---	---	--------	--------	--------	--------

Odds Ratio Estimates

Point 95% Wald

Effect	ICE_CREAM	Estimate	w	Confidence Limits	x
--------	-----------	----------	---	-------------------	---

VIDEO	1	0.955	0.909	1.003
-------	---	-------	-------	-------

VIDEO	2	0.977	0.938	1.018
-------	---	-------	-------	-------

PUZZLE	1	0.921	0.879	0.965
--------	---	-------	-------	-------

PUZZLE	2	0.958	0.921	0.996
--------	---	-------	-------	-------

FEMALE	1	2.338	0.971	5.628
--------	---	-------	-------	-------

FEMALE 2 1.033 0.520 2.052

p. Parameter - This column lists the predictor values and the intercept-the parameters that were estimated in the model. The intercept and each predictor appears twice because two models were fitted.

q. ICE_CREAM - Two models were defined in this multinomial regression: one relating chocolate to the referent category, strawberry, and another model relating vanilla to strawberry. The ice_cream number indicates to which model an estimate, standard error, chi-square, and p-value refer. We can refer to the response profiles to determine which response corresponds to which model. Our ice_cream categories 1 and 2 are chocolate and vanilla, respectively, so values of 1 correspond to the chocolate relative to strawberry model and values of 2 correspond to the

vanilla relative to strawberry model.

r. DF - These are the degrees of freedom for parameter in the specified model. Since our predictors are continuous variables, they all have one degree of freedom in each model.

s.

Estimate -

These are the estimated multinomial logistic regression coefficients for the models. An important feature of the multinomial logit model is that it estimates $k-1$ models, where k is the number of levels of the outcome variable. SAS treats strawberry as the referent group and estimates a model for chocolate relative to strawberry and a model for vanilla relative to strawberry. Therefore, each estimate listed in this column must be considered in terms both the parameter it corresponds to and the model to which it belongs. The standard interpretation of the

multinomial logit is that for a unit change in the predictor variable, the logit of outcome m relative to the referent group is expected to change by its respective parameter estimate (which is in log-odds units) given the other variables in the model are held constant.

Model Number 1: chocolate relative to strawberry

Intercept - This is the multinomial logit estimate for chocolate relative to strawberry when the predictor variables in the model are evaluated at zero. For males (the variable female evaluated at zero) with zero

video and puzzle scores, the logit for preferring chocolate to strawberry is 5.9696. Note that evaluating video and puzzle at

zero is out of the range of plausible scores. If the scores were mean-centered, the intercept would have a natural interpretation: log odds of preferring chocolate to strawberry for a male with average video and puzzle scores.

video - This is the multinomial logit estimate for a one unit increase in video score for chocolate relative to strawberry, given the other variables in the model are held constant. If a subject were to increase his

video score by one point, the multinomial log-odds for preferring chocolate to strawberry would be expected to decrease by 0.0465 unit while holding all other variables in the model constant.

puzzle - This is the multinomial logit estimate for a one unit

increase in puzzle score for chocolate relative to strawberry, given the other variables in the model are held constant. If a subject were to increase his puzzle score by one point, the multinomial log-odds for preferring chocolate to strawberry would be expected to decrease by 0.0819 unit while holding all other variables in the model constant.

female - This is the multinomial logit estimate comparing females to males for chocolate relative to strawberry, given the other variables in the model are held constant. The multinomial logit for females relative to males is 0.8495 unit higher for preferring chocolate to strawberry, given all other predictor variables in the model are held constant. In other words, females are more likely than males to prefer chocolate to strawberry.

Model 2: vanilla relative to strawberry

Intercept - This is the multinomial logit estimate for vanilla relative to strawberry when the other predictor variables in the model are evaluated at zero. For males (the variable female evaluated at zero) with zero video and puzzle scores, the logit for preferring vanilla to strawberry is 4.0572.

video - This is the multinomial logit estimate for a one unit increase in video score for vanilla relative to strawberry, given the other variables in the model are held constant. If a subject were to increase his

video score by one point, the multinomial log-odds for preferring vanilla to strawberry would be expected to decrease by 0.0229 unit while holding all other variables in the model constant.

puzzle - This is the multinomial logit estimate for a one

unit

increase in puzzle score for vanilla relative to strawberry, given the other variables in the model are held constant. If a subject were to increase his puzzle score by one point, the multinomial log-odds for preferring vanilla to strawberry would be expected to decrease by 0.0430 unit while holding all other variables in the model constant.

female - This is the multinomial logit estimate comparing females to males for vanilla relative to strawberry, given the other variables in the model are held constant. The multinomial logit for females relative to males is 0.0328 unit higher for preferring vanilla to strawberry, given all other predictor variables in the model are held constant. In other words, males are less likely than females to prefer vanilla ice cream to strawberry ice cream.

t.

Standard Error - These are the standard errors of the individual regression coefficients for the two respective models estimated.

u.

Chi-Square -

This column lists the Chi-Square test statistic of the given parameter and model.

v.

Pr > Chi-Square - This is the p-value used to determine whether or

not the null hypothesis that a particular predictor's regression coefficient is

zero, given that the rest of the predictors are in the model, can be rejected.

If the p-value less than alpha, then the null hypothesis can be rejected and the

parameter estimate is considered to be statistically significant at that alpha

level. The Chi-Square

test statistic values follows a Chi-Square

distribution which is used to test against the alternative hypothesis that the estimate is not equal to zero. In multinomial logistic regression, the interpretation of a parameter estimate's significance is limited to the model in which the parameter estimate was calculated. For example, the significance of a parameter estimate in the chocolate relative to strawberry model cannot be assumed to hold in the vanilla relative to strawberry model.

Model 1: chocolate relative to strawberry

For chocolate relative to strawberry, the Chi-Square test statistic for the intercept is 17.2425 with an associated p-value of <0.0001 . With an alpha level of 0.05, we would reject the null hypothesis and conclude that the multinomial logit for males (the variable female evaluated at zero) and

with zero video and puzzle scores in chocolate relative to strawberry are found to be statistically different from zero.

For chocolate relative to strawberry, the Chi-Square test statistic for the predictor video is 3.4296 with an associated p-value of 0.0640. If we set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that for chocolate relative to strawberry, the regression coefficient for video has not been found to be statistically different from zero given puzzle and female are in the model.

For chocolate relative to strawberry, the Chi-Square test statistic for the predictor puzzle is 11.8149 with an associated p-value of 0.0006. If we again set our alpha level to 0.05, we would reject the null hypothesis and

conclude that the regression coefficient for puzzle has been found to be statistically different from zero for chocolate relative to strawberry given that video and female are in the model.

For chocolate relative to strawberry, the Chi-Square test statistic for the predictor female is 3.5913 with an associated p-value of 0.0581. If we again set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that the difference between males and females has not been found to be statistically different for chocolate relative to strawberry given that video and puzzle are in the model.

Model 2: vanilla relative to strawberry

For vanilla relative to strawberry, the Chi-Square test statistic for the intercept is 11.0065 with an associated p-value of

0.0009. With an alpha level of 0.05, we would reject the null hypothesis and conclude that a) the multinomial logit for males (the variable female evaluated at zero) and with zero

video and

puzzle scores in vanilla relative to strawberry are statistically different from zero; or b) for males with zero video and

puzzle scores, there is a statistically significant difference between the likelihood of being classified as preferring vanilla or preferring strawberry.

Such a male would be more likely to be classified as preferring vanilla to

strawberry. We can make the second interpretation when we view the intercept as a specific covariate profile (males with zero

video and

puzzle

scores). Based on the direction and significance of the coefficient, the

intercept

indicates whether the profile would have a greater propensity to be classified in one level of the outcome variable than the other level.

For vanillarelated to strawberry, the Chi-Square test statistic for the predictor video is 1.2060 with an associated p-value of 0.2721. If we set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that for vanilla relative to strawberry, the regression coefficient for video has not been found to be statistically different from zero given

puzzle and female are in the model.

For vanilla relative to strawberry, the Chi-Square test statistic for the predictor puzzle is 4.6746 with an associated p-value of 0.0306. If we again set our alpha level to 0.05, we would reject the

null hypothesis and conclude that the regression coefficient for puzzle has been found to be statistically different from zero for vanilla relative to strawberry given that video and female are in the model.

For vanilla relative to strawberry, the Chi-Square test statistic for the predictor female is 0.0088 with an associated p-value of 0.9252. If we again set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that for vanilla relative to strawberry, the regression coefficient for female has not been found to be statistically different from zero given puzzle and video are in the model.

w. Odds Ratio Point Estimate - These are the proportional odds ratios.

They can be obtained by exponentiating the estimate,

eestimate.

x. 95% Wald Confidence Limits - This is the Confidence Interval (CI)

for the proportional odds ratio given the other predictors are in the model. For

a given predictor with a level of 95% confidence, we say that we are 95%

confident that the "true" population proportional odds ratio lies between the

lower and upper limit of the interval. The CI is equivalent to the Wald

Chi-Square test statistic; if the CI includes 1, we would fail to reject the

null hypothesis that a particular ordered logit regression coefficient is zero

given the other predictors are in the model at an alpha level of 0.05. The CI is

more illustrative than the Wald Chi-Square test statistic.