

How do you perform Multinomial Logistic Regression using Stata, and what does the annotated output show?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you perform Multinomial Logistic Regression using Stata, and what does the annotated output show?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160017>

Multinomial logistic regression is a statistical method used to analyze the relationship between a categorical dependent variable with more than two categories and one or more independent variables. This method is commonly used in social science research and can be performed using the statistical software Stata.

To perform multinomial logistic regression in Stata, one must first load the data and specify the dependent variable and independent variables. Then, the "mlogit" command is used to estimate the model and generate the output. The output includes various statistical measures such as the coefficients, standard errors, and p-values, which indicate the significance of each independent variable in predicting the dependent variable.

The annotated output in Stata provides a detailed summary of the results, including the model's goodness of fit, the significance of the overall model, and the classification table. Additionally, the output also includes the odds ratio for each independent variable, which indicates the likelihood of a particular category of the dependent variable occurring relative to the reference category.

Overall, the annotated output in Stata provides valuable information for interpreting the results of multinomial logistic regression, allowing researchers to identify the significant predictors and understand the relationship between the dependent and independent variables.

Multinomial Logistic Regression | Stata Annotated Output

This page shows an example of a multinomial logistic regression analysis with footnotes explaining the output. The data were 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 exist 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 Stata to choose the referent group. In our example, this will be vanilla. By default, Stata chooses the most frequently occurring group to be the referent group. The first half of this page interprets the coefficients in terms of multinomial log-odds (logits). These will be close to but not equal to the log-odds achieved in a logistic regression with two levels of the outcome variable. The second half interprets the coefficients in terms of relative risk ratios.

use `https://stats.idre.ucla.edu/stat/stata/output/mlogit, clear`

Before running the regression, obtaining a frequency of the ice cream flavors in the data can inform the selection of a reference group.

tab ice_cream

**favorite flavor of ice cream
| Freq. Percent Cum.**

```
-----+-----  
chocolate | 47 23.50 23.50  
vanilla | 95 47.50 71.00  
strawberry | 58 29.00 100.00  
-----+-----  
Total | 200 100.00
```

Vanilla is the most frequently occurring ice cream flavor and will be the reference group in this example.

mlogit ice_cream video puzzle female

Iteration 0: log likelihood = -210.58254

Iteration 1: log likelihood = -194.75041

Iteration 2: log likelihood = -194.03782

Iteration 3: log likelihood = -194.03485

Iteration 4: log likelihood = -194.03485

Multinomial logistic regression Number of obs = 200

LR chi2(6) = 33.10

Prob > chi2 = 0.0000

Log likelihood = -194.03485 Pseudo R2 = 0.0786

ice_cream | Coef. Std. Err. z P>|z|
-----+

chocolate |

video | -.0235647 .0209747 -1.12 0.261 -.0646744 .017545

**puzzle | -.0389243 .0195165 -1.99 0.046 -.0771759 -
.0006726**

female | .8166202 .3909813 2.09 0.037 .050311 1.582929

_cons | 1.912256 1.127256 1.70 0.090 -.2971258 4.121638
-----+

strawberry |

video | .022922 .0208718 1.10 0.272 -.0179861 .0638301

puzzle | .0430036 .0198894 2.16 0.031 .0040211 .081986

**female | -.032862 .3500153 -0.09 0.925 -.7188793
.6531553**

_cons | -4.057323 1.222939 -3.32 0.001 -6.45424

-1.660407

(ice_cream==vanilla is the base outcome)

Iteration Loga

Iteration 0: log likelihood = -210.58254

Iteration 1: log likelihood = -194.75041

Iteration 2: log likelihood = -194.03782

Iteration 3: log likelihood = -194.03485

Iteration 4: log likelihood = -194.03485

a. Iteration Log - This is a listing of the log likelihoods at each

iteration. Remember that multinomial logistic regression, like binary and

ordered logistic regression, uses maximum likelihood estimation, which is an

iterative procedure. The first iteration (called iteration 0) is the log

likelihood of the "null" or "empty" model; that is, a model with no predictors.

At the next iteration, the predictor(s) are included in the model. At each

iteration, the log likelihood increases because the goal is to maximize the log likelihood. When the difference between successive iterations is very small, the model is said to have "converged", the iterating stops, and the results are displayed. For more information on this process for binary outcomes, see

Regression Models for Categorical and Limited Dependent Variables by J. Scott Long (page 52-61).

Model Summary

Multinomial logistic regression Number of obs = 200

LR chi2(6) = 33.10

Prob > chi2 = 0.0000

Log likelihood = -194.03485 Pseudo R2 = 0.0786

b. Log Likelihood - This is the log likelihood of the fitted model. It

is used in the Likelihood Ratio Chi-Square test of whether all predictors'

regression coefficients in the model are simultaneously

zero and in tests of nested models.

c. Number of obs - This is the number of observations used in the multinomial logistic regression. It may be less than the number of cases in the dataset if there are missing values for some variables in the equation. By default, Stata does a listwise deletion of incomplete cases.

d. LR chi2(6) - This is the Likelihood Ratio (LR) Chi-Square test that for both equations (chocolate relative to vanilla and strawberry relative to vanilla) that at least one of the predictors' regression coefficient is not equal to zero. The number in the parentheses indicates the degrees of freedom of the Chi-Square distribution used to test the LR Chi-Square statistic and is defined by the number of models estimated (2) times the number of predictors in

the model (3). The LR Chi-Square statistic can be calculated by $-2 * (L(\text{null model}) - L(\text{fitted model})) = -2 * ((-210.583) - (-194.035)) = 33.096$, where $L(\text{null model})$ is from the log likelihood with just the response variable in the model (Iteration 0) and $L(\text{fitted model})$ is the log likelihood from the final iteration (assuming the model converged) with all the parameters.

e. Prob > chi2 - This is the probability of getting a LR test statistic as extreme as, or more so, than the observed statistic under the null hypothesis; the null hypothesis is that all of the regression coefficients across both models are simultaneously equal to zero. In other words, this is the probability of obtaining this chi-square statistic (33.10) or one more extreme if there is in fact no effect of the predictor variables. This p-value is compared to a specified alpha level, our willingness to

accept a type I error, which is typically set at 0.05 or 0.01. The small p-value from the LR test, <0.00001 , would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero. The parameter of the chi-square distribution used to test the null hypothesis is defined by the degrees of freedom in the prior line, $\text{chi2}(6)$.

f. Pseudo R² - This is McFadden's pseudo R-squared. Logistic regression does not have an equivalent to the R-squared that is found in OLS regression; however, many people have tried to come up with one. There are a wide variety of pseudo-R-square statistics. Because this statistic does not mean what R-square means in OLS regression (the proportion of variance of the response variable explained by the predictors), we suggest interpreting this statistic with great caution.

Parameter Estimates

ice_creamg | Coef. h Std. Err. z k P>|z|k l
-----+

chocolate |

video | -.0235647 .0209747 -1.12 0.261 -.0646744 .017545
puzzle | -.0389243 .0195165 -1.99 0.046 -.0771759 -
.0006726
female | .8166202 .3909813 2.09 0.037 .050311 1.582929
_cons | 1.912256 1.127256 1.70 0.090 -.2971258 4.121638

-----+
strawberry |

video | .022922 .0208718 1.10 0.272 -.0179861 .0638301
puzzle | .0430036 .0198894 2.16 0.031 .0040211 .081986
female | -.032862 .3500153 -0.09 0.925 -.7188793
.6531553
_cons | -4.057323 1.222939 -3.32 0.001 -6.45424
-1.660407

(ice_cream==vanilla is the base outcome)i

g. ice_cream - This is the response variable in the multinomial

logistic regression. Underneath ice_cream are two replicates of the predictor variables, representing the two models that are estimated: chocolate relative to vanilla and strawberry relative to vanilla.

h and i. Coef. and referent group - These are the estimated multinomial logistic regression coefficients and the referent level, respectively, for the model. 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. In this instance, Stata, by default, set vanilla as the referent group, and therefore estimated a model for chocolate relative to vanilla and a model for strawberry relative to vanilla. Since the parameter estimates are relative to the referent group, 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 variables in the model are held constant.

chocolate relative to vanilla

video - This is the multinomial logit estimate for a one unit increase in video score for chocolate relative to vanilla, 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 vanilla would be expected to decrease by 0.024 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 vanilla, 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 vanilla would be expected to decrease by 0.039 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 vanilla, given the other variables in the model are held constant. The multinomial logit for females relative to males is 0.817 unit higher for preferring chocolate to vanilla, given all other predictor variables in the model are held constant. In other words, females are more likely than males to prefer chocolate to vanilla.

_cons - This is the multinomial logit estimate for chocolate relative to vanilla 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

vanilla is 1.912. 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 vanilla for a male with average video and puzzle scores.

strawberry relative to vanilla

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

increase in video score for strawberry relative to vanilla, 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

strawberry to vanilla would be expected to increase by 0.023 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 strawberry relative to vanilla, 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

strawberry to vanilla would be expected to increase by 0.043 unit while holding

all other variables in the model constant.

female - This is the multinomial logit estimate comparing females

to males for strawberry relative to vanilla, given the other variables in the

model are held constant. The multinomial logit for females relative to males is

0.033 unit lower for preferring strawberry to vanilla, given all other predictor

variables in the model are held constant. In other words, males are more likely

than females to prefer strawberry ice cream to vanilla ice cream.

_cons - This is the multinomial logit estimate for strawberry relative to vanilla 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 strawberry to vanilla is -4.057.

j. Std. Err. - These are the standard errors of the individual regression coefficients for the two respective models estimated. They are used in both the calculation of the z test statistic, superscript k, and the confidence interval of the regression coefficient, superscript l.

k. z and $P > |z|$ - The test statistic z is the ratio of the Coef. to the Std. Err. of the respective predictor, and the

p-value $P > |z|$ is the probability the z test statistic (or a more extreme test statistic) would be observed under the null hypothesis. For a given alpha level, z and $P > |z|$ 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 $P > |z|$ is less than alpha, then the null hypothesis can be rejected and the parameter estimate is considered significant at that alpha level. The z value follows a standard normal distribution which is used to test against a two-sided alternative hypothesis that the Coef. 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 vanilla model cannot be assumed to hold in the strawberry relative to vanilla model.

chocolate relative to vanilla

For chocolate relative to vanilla, the z test statistic for the

predictor video $(-0.024/0.021)$ is -1.12 with an associated p-value of

0.261 . If we set our alpha level to 0.05 , we would fail to reject the null

hypothesis and conclude that for chocolate relative to vanilla, 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 vanilla, the z test statistic for the predictor puzzle $(-0.039/0.020)$ is -1.99 with an associated p-value

of 0.046 . 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 chocolaterelative to vanilla given that video and female are in the model. For chocolaterelative to vanilla, the z test statistic for the predictor female (0.817/0.391) is 2.09 with an associated p-value of 0.037. If we again set our alpha level to 0.05, we would reject the null hypothesis and conclude that the difference between males and females has been found to be statistically different for chocolate relative to vanilla given that video and female are in the model. For chocolaterelative to vanilla, the z test statistic for the intercept, _cons (1.912/1.127) is 1.70 with an associated p-value of 0.090. With an alpha level of 0.05, we would fail to 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 chocolaterelative to vanilla are found not to be statistically different

from zero; or b) for males with zero video and puzzle scores, you are statistically uncertain whether they are more likely to be classified as preferring chocolate or vanilla. We can make the second interpretation when we view the `_cons` as a specific covariate profile (males with zero video and puzzle scores). Based on the direction and significance of the coefficient, the `_cons` indicates whether the profile would have a greater propensity to be classified in one level of the outcome variable than the other level.

strawberry relative to vanilla

For strawberry relative to vanilla, the z test statistic for the predictor video ($0.023/0.021$) is 1.10 with an associated p-value of 0.272. If we set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that for strawberry relative to

vanilla, the regression coefficient for video has not been found to be statistically different from zero given puzzle and female are in the model. For strawberry relative to vanilla, the z test statistic for the predictor puzzle (0.043/0.020) is 2.16 with an associated p-value of 0.031. 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 strawberry relative to vanilla given that video and female are in the model.

For strawberry relative to vanilla, the z test statistic for the predictor female (-0.033/0.350) is -0.09 with an associated p-value of 0.925. If we again set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that for strawberry relative to

vanilla, the regression coefficient for female has not been found to be statistically different from zero given puzzle and video are in the model. For strawberry relative to vanilla, the z test statistic for the intercept, `_cons` (-4.057/1.223) is -3.32 with an associated p-value of 0.001. 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 strawberry relative to vanilla 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 strawberry or preferring vanilla. 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 `_cons` as a specific covariate profile (males with zero video and puzzle scores). Based on the direction and significance of the coefficient, the `_cons` indicates whether the profile would have a greater propensity to be classified in one level of the outcome variable than the other level.

I. - This is the Confidence Interval (CI) for an individual multinomial logit regression coefficient given the other predictors are in the model for outcome m relative to the referent group. For a given predictor with a level of 95% confidence, we'd say that we are 95% confident that the "true" population multinomial logit regression coefficient lies between the lower and upper limit of the interval for outcome m relative to the referent group. It is calculated as the $\text{Coef. } (z_{\alpha/2}) * (\text{Std.Err.})$, where $z_{\alpha/2}$ is a critical value on the standard normal

distribution.

The CI is equivalent to the z test statistic: if the CI includes zero, we'd fail to reject the null hypothesis that a particular regression coefficient is zero given the other predictors are in the model. An advantage of a CI is that it is illustrative; it provides a range where the "true" parameter may lie.

Relative Risk Ratio Interpretation

The following is the interpretation of the multinomial logistic regression in terms of relative risk ratios and can be obtained by mlogit, rrr after running the multinomial logit model or by specifying the rrr option when the full model is specified. This part of the interpretation applies to the output below.

mlogit ice_cream video puzzle female, rrr

Iteration 0: log likelihood = -210.58254

Iteration 1: log likelihood = -194.75041

Iteration 2: log likelihood = -194.03782

Iteration 3: log likelihood = -194.03485

Iteration 4: log likelihood = -194.03485

Multinomial logistic regression Number of obs = 200

LR chi2(6) = 33.10

Prob > chi2 = 0.0000

Log likelihood = -194.03485 Pseudo R2 = 0.0786

ice_cream | RRRa Std. Err. z P>|z| b
-----+

chocolate |

video | .9767108 .0204862 -1.12 0.261 .9373726 1.0177

puzzle | .9618236 .0187714 -1.99 0.046 .925727 .9993276

female | 2.262839 .8847276 2.09 0.037 1.051598 4.869199
-----+

strawberry |

video | 1.023187 .0213558 1.10 0.272 .9821747 1.065911

puzzle | 1.043942 .0207633 2.16 0.031 1.004029 1.085441

female | .9676721 .3387 -0.09 0.925 .4872981 1.921595

(ice_cream==vanilla is the base outcome)

a. Relative Risk Ratio - These are the relative risk ratios for the multinomial logit model shown earlier. They can be obtained by exponentiating the multinomial logit coefficients, `ecoef`, or by specifying the `rrr` option when the `mlogit` command is issued. Recall that the multinomial logit model estimates $k-1$ models, where the k th equation is relative to the referent group. The RRR of a coefficient indicates how the risk of the outcome falling in the comparison group compared to the risk of the outcome falling in the referent group changes with the variable in question. An $RRR > 1$ indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group increases as the variable increases. In other words, the comparison outcome is more likely. An

RRR < 1

indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group decreases as the variable increases. See the interpretations of the relative risk ratios below for examples. In general, if the $RRR < 1$, the outcome is more likely to be in the referent group.

chocolate relative to vanilla

video - This is the relative risk ratio for a one unit increase in video score for preferring chocolate to vanilla, given that the other variables in the model are held constant. If a subject were to increase her video score by one unit, the relative risk for preferring chocolate to vanilla would be expected to decrease by a factor of 0.977 given the other variables in the model are held constant. So, given a

one unit increase in video, the relative risk of being in the chocolate group would be 0.977 times more likely when the other variables in the model are held constant. More generally, we can say that if a subject were to increase her video score, we would expect her to be more likely to prefer vanilla ice cream over chocolate ice cream.

puzzle - This is the relative risk ratio for a one unit increase in puzzle score for preferring chocolate to vanilla, given that the other variables in the model are held constant. If a subject were to increase her puzzle score by one unit, the relative risk for preferring chocolate to vanilla would be expected to decrease by a factor of 0.962 given the other variables in the model are held constant. More generally, we can say that if two subjects have identical video scores and are both

**female (or both male),
the subject with the higher puzzle score is more likely
to prefer vanilla
ice cream over chocolate ice cream than the subject
with the lower puzzle
score.**

**female - This is the relative risk ratio comparing females
to
males for preferring chocolate to vanilla, given that the
other variables in the
model are held constant. For females relative to males,
the relative risk for
preferring chocolate relative to vanilla would be
expected to increase by a
factor of 2.263 given the other variables in the model
are held constant. In
other words, females are more likely than males to
prefer chocolate ice cream
over vanilla ice cream.**

strawberry relative to vanilla

**video - This is the relative risk ratio for a one unit
increase in**

video score for preferring strawberry to vanilla, given that the other variables in the model are held constant. If a subject were to increase her video score by one unit, the relative risk for strawberry relative to vanilla would be expected to increase by a factor of 1.023 given the other variables in the model are held constant. More generally, we can say that if a subject were to increase her video score, we would expect her to be more likely to prefer strawberry ice cream over vanilla ice cream.

puzzle - This is the relative risk ratio for a one unit increase in puzzle score for preferring strawberry to vanilla, given that the other variables in the model are held constant. If a subject were to increase her puzzle score by one unit, the relative risk for strawberry relative to vanilla would be expected to increase by a factor of 1.043 given the

other variables in the model are held constant. More generally, we can say that if two subjects have identical video scores and are both female (or both male), the subject with the higher puzzle score is more likely to prefer strawberry ice cream to vanilla ice cream than the subject with the lower puzzle score.

female - This is the relative risk ratio comparing females to males for strawberry relative to vanilla, given that the other variables in the model are held constant. For females relative to males, the relative risk for preferring strawberry to vanilla would be expected to decrease by a factor of 0.968 given the other variables in the model are held constant. In other words, females are less likely than males to prefer strawberry ice cream to vanilla ice cream.

b. - This is the CI for the relative risk ratio given the other predictors are in the model. For a given predictor with a level of 95% confidence, we'd say that we are 95% confident that the "true" population relative risk ratio comparing outcome m to the referent group lies between the lower and upper limit of the interval. An advantage of a CI is that it is illustrative; it provides a range where the "true" relative risk ratio may lie.