What is Ordinal Logistic Regression and how is it used in SAS Data Analysis?

Authored by

stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is Ordinal Logistic Regression and how is it used in SAS Data Analysis?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=157794

Ordinal Logistic Regression is a statistical technique used in SAS Data Analysis to predict the probability of an event occurring based on a set of independent variables. It is primarily used when the dependent variable is ordinal in nature, meaning it has a specific order or ranking, and cannot be categorized into discrete groups. This technique uses the cumulative probability of the event occurring to estimate the relationship between the independent variables and the likelihood of the event occurring. It is often used in social sciences, marketing, and medical research to analyze data and make predictions about the likelihood of an event happening on a scale or in a specific order. Ordinal Logistic Regression is a valuable tool in SAS Data Analysis for understanding and predicting outcomes in scenarios where the dependent variable has a natural ordering.

Ordinal Logistic Regression | SAS Data Analysis Examples

Version info: Code for this page was tested in SAS 9.3.

Examples of ordered logistic regression

Example 1: A marketing research firm wants to investigate what factors influence the size of soda (small, medium, large or extra large) that people order at a fast-food chain. These factors may include what type of sandwich is ordered (burger or chicken), whether or not fries are also ordered, and age of the consumer. While the outcome variable, size of soda, is obviously ordered, the difference between the various sizes is not consistent.

The differences are 10, 8, 12 ounces, respectively.

Example 2: A researcher is interested in what factors influence medaling in Olympic swimming. Relevant predictors include at training hours, diet, age, and popularity of swimming in the athlete's home country. The researcher believes that the distance between gold and silver is larger than the distance between silver and bronze.

Example 3: A study looks at factors that influence the decision of whether to apply to graduate school. College juniors are asked if they are unlikely, somewhat likely, or very likely to apply to graduate school. Hence, our outcome variable has three categories. Data on parental educational status, whether the undergraduate institution is public or private, and current GPA is also collected. The researchers have reason to believe that the "distances" between these three

points are not equal. For example, the "distance" between "unlikely" and "somewhat likely" may be shorter than the distance between "somewhat likely" and "very likely".

Description of the data

For our data analysis below, we are going to expand on Example 3 about applying to graduate school. We have generated hypothetical data, which can be downloaded: ologit.

This hypothetical data set has a three-level variable called apply (coded 0, 1, 2), that we will use as our response (i.e., outcome, dependent) variable. We also have three variables that we will use as predictors: pared, which is a 0/1 variable indicating whether at least one parent has a graduate degree; public, which is a 0/1 variable where 1 indicates that the undergraduate institution is a public university

and 0 indicates that it is

a private university, and gpa, which is the student's grade point average.

proc freq data = ologit;

tables apply;

tables pared;

tables public;

run;

APPLY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	220	55.00	220	55.00
1	140	35.00	360	90.00
2	40	10.00	400	100.00

PARED	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	337	84.25	337	84.25
1	63	15.75	400	100.00

PUBLIC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	343	85.75	343	85.75
1	57	14.25	400	100.00

proc means data = ologit;

var gpa;

run;

The MEANS Procedure

Analysis Variable : GPA

N Mean Std Dev Minimum Maximum

----

400 2.9989250 0.3979409 1.9000000 4.0000000

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.

Ordered logistic regression

Before we run our ordinal logistic model, we will see if any cells (created
by the crosstab of our categorical and response variables) are empty or
extremely small. If any are, we may have difficulty running our model.
We have used some options on the tables statements to

clean up the output.

Perhaps the most important option is the missprint option; this will have
SAS include missing values as a category in the table. Because we have no
missing values in this data set, this option is not really needed; we have
included it here only to show its use.

```
proc freq data = ologit;
tables apply*pared / nopercent norow nocol missprint;
tables apply*public / nopercent norow nocol missprint;
run;
```

The FREQ Procedure

Table of APPLY by PARED

APPLY PARED

```
Frequency| 0| 1| Total
---------+--------+--------+
0 | 200 | 20 | 220
---------+--------+--------+
1 | 110 | 30 | 140
```

---------+--------+--------+

2 | 27 | 13 | 40

---------+--------+--------+

Total 337 63 400

Table of APPLY by PUBLIC

APPLY PUBLIC

Frequency| 0| 1| Total

---------+--------+--------+

0 | 189 | 31 | 220

---------+--------+--------+

1 | 124 | 16 | 140

---------+--------+--------+

2 | 30 | 10 | 40

---------+--------+--------+

Total 343 57 400

None of the cells is too small or empty (has no cases), so we will run our
model.

proc logistic data = ologit desc;
class pared(ref='0') public(ref='0') / param=reference;

model apply = pared public gpa;

run;

The LOGISTIC Procedure

Model Information

Data Set ologit Written by SAS

Response Variable APPLY

Number of Response Levels 3

Model cumulative logit

Optimization Technique Fisher's scoring

Number of Observations Read 400

Number of Observations Used 400

Response Profile

Ordered Total

Value APPLY Frequency

1 2 40

2 1 140

3 0 220

Probabilities modeled are cumulated over the lower

Ordered Values.

Model Convergence Status

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

Score Test for the Proportional Odds Assumption

Chi-Square DF Pr > ChiSq

4.8446 3 0.1835

Model Fit Statistics

Intercept
Intercept and
Criterion Only Covariates

AIC 745.205 727.025
SC 753.188 746.982
-2 Log L 741.205 717.025

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test Chi-Square DF Pr > ChiSq

Likelihood Ratio 24.1804 3 <.0001

Score 23.4804 3 <.0001

Wald 24.3337 3 <.0001

Analysis of Maximum Likelihood Estimates

Standard Wald

Parameter DF Estimate Error Chi-Square Pr > ChiSq

Intercept 2 1 -4.2983 0.8092 28.2189 <.0001

Intercept 1 1 -2.2029 0.7844 7.8869 0.0050

PARED 1 1.0478 0.2684 15.2350 <.0001

PUBLIC 1 -0.0585 0.2886 0.0411 0.8393

GPA 1 0.6156 0.2626 5.4963 0.0191

Odds Ratio Estimates

Point 95% Wald

Effect Estimate Confidence Limits

PARED 2.851 1.685 4.826

PUBLIC 0.943 0.536 1.661

GPA 1.851 1.106 3.096

Association of Predicted Probabilities and Observed Responses

Percent Concordant 60.0 Somers' D 0.210
Percent Discordant 39.0 Gamma 0.213
Percent Tied 1.1 Tau-a 0.119
Pairs 45200 c 0.605

In the output above, we see that all 400 observations in our data set were used in the analysis. Fewer observations would have been used if any of our variables had missing values. By default, SAS does a listwise deletion of cases with missing values. The Response Profile shows the value that SAS used when conducting the analysis (given in the Ordered Value column), the value of the original variable, and the number of cases in each level of the outcome variable. (If you want SAS to use the values that you have assigned the outcome variable, then you would want to use the order = data option on the proc logistic statement.) The note below this table reminds us that the "Probabilities modeled are cumulated over the lower

Ordered Values."

It is helpful to remember this when interpreting the output. Next we see

that the model converged (you should not try to interpret any output if the

model has not converged), and we also see that the test of the proportional odds

assumption is non-significant. One of the assumptions underlying ordinal

logistic (and ordinal probit) regression is that the relationship between each

pair of outcome groups is the same. In other words, ordinal logistic

regression assumes that the coefficients that describe the relationship between,

say, the lowest versus all higher categories of the response variable are the

same as those that describe the relationship between the next lowest category

and all higher categories, etc. This is called the proportional odds

assumption or the parallel regression assumption. Because the relationship

between all pairs of groups is the same, there is only

one set of coefficients
(only one model). If this was not the case, we would need different models
(such as a generalized ordered logit model) to describe the relationship between
each pair of outcome groups. The table showing the Model Fit Statistics provides the AIC, SC and -2 log likelihood. These can be used in the comparison of nested models. In
the next table we see various tests of the overall model; they all indicated
that the model is statistically significant.

In the table Analysis of Maximum Likelihood Estimates, we see the degrees of
freedom, coefficients, their standard errors, the Wald chi-square test and
associated p-values.
Both pared and gpa are statistically significant; public is
not. So for pared, we would say that for a one unit increase in pared (i.e., going from 0 to 1), we expect a 1.05 increase in
the log odds of being in a higher level of apply, given all

of the other variables in the model are held constant. For gpa, we would say that for a one unit increase in gpa, we would expect a 0.62 increase in the log odds of being in a higher level of apply, given that all of the other variables in the model are held constant. In the next table we see the results presented as proportional odds ratios (the coefficient exponentiated) and the 95% confidence intervals for the proportional odds ratios. We would interpret the proportional odds ratios pretty much as we would odds ratios from a binary logistic regression. For pared, we would say that for a one unit increase in pared, i.e., going from 0 to 1, the odds of high apply versus the combined middle and low categories are 2.85 greater, given that all of the other variables in the model are held constant. Likewise, the odds of the combined middle and high categories versus low apply

is 2.85 times greater,
given that all of the other variables in the model are held constant. For a one unit
increase in gpa, the odds of the high category of apply versus the low and middle categories of apply are 1.85 times greater, given that the
other variables in the model are held constant. Because of the
proportional odds assumption (see below for more explanation), the same
increase, 1.85 times, is found between low apply and the combined
categories of middle and high apply.

We can also obtain predicted probabilities, which are usually easier to
understand than the coefficients or the odds ratios. We will use the
estimate statement. To use the estimate statement, we supply
values of our predictor variables to be multiplied by the regression
coefficients, which are for our current model the intercept for apply =

**2, the intercept for apply = 1, the coefficient for public = 1 ,
the coefficient for pared = 1, and the coefficient for gpa. Here we will
see how the probabilities of membership to the categories of apply change
as we vary pared and hold public at 1 and gpa at its mean
of 2.9989.**

```
proc logistic data = ologit desc;;
class pared(ref='0') public(ref='0')/ param = reference;
model apply = pared public gpa;
estimate "Pr prob apply=2 at pared=0" intercept 1 public 1 gpa 2.9989 / ilink category='2';
estimate "Pr prob apply=2 at pared=1" intercept 1 pared 1 public 1 gpa 2.9989 / ilink category='2';
estimate "Pr prob apply=1 or 2 at pared=0" intercept 1 public 1 gpa 2.9989 / ilink category='1';
estimate "Pr prob apply=1 or 2 at pared=1" intercept 1 pared 1 public 1 gpa 2.9989 / ilink category='1';
run;
```

*****SOME OUTPUT OMITTED AND LAYOUT MODIFIED*****

Label APPLY Estimate Standard Error z Value Pr > |z| Mean Standard Error
of Mean
Pr prob apply=2 at pared=0 2 -2.5108 0.3104 -8.09 <.0001 0.07511 0.02156
Pr prob apply=2 at pared=1 2 -1.4629 0.3545 -4.13 <.0001 0.188 0.05412
Pr prob apply=1 or 2 at pared=0 1 -0.4153 0.2733 -1.52 0.1286 0.3976 0.06546
Pr prob apply=1 or 2 at pared=1 1 0.6325 0.3451 1.83 0.0668 0.6531 0.07818

The predicted probabilities are listed in the "Mean" column. All
predicted probabilities discussed below were calculated at public = 1 and
gpa = 2.9989. As you can see, the predicted probability of
being in the highest category of apply (apply = 2) is 0.07511 if neither parent has a graduate
level education and 0.1880 otherwise. For membership to either the
highest or middle category of apply (apply = 1 or 2), the predicted probabilities are 0.3976 and 0.6531, for

parents without graduate level education and with graduate level education, respectively. Predicted probabilities of being in the middle category alone can be calculated by subtracting the predicted probabilities of (apply = 1 or 2) from the probability of (apply = 2). Thus, the probability of belonging to the middle apply category when parents do not have graduate level education is 0.3976 - 0.07511 = 0.32249. Predicted probabilities of being in the lowest apply category can be obtained in 2 ways. First, we can subtract the probability of being in either the highest or middle apply category from 1. For example, the probability of being in the lowest apply group (apply = 0) when parents do not have graduate education is 1 - 0.3976 = 0.6024. Alternatively, we can change the reference apply category to 2 by removing the desc option from the proclogistic

statement and supply a new estimate statement to get the probabilities of being in apply category 0.

```
proc logistic data = ologit;
class pared(ref='0') public(ref='0')/ param = reference;
model apply = pared public gpa;
estimate "Pr prob apply=0 at pared=0" intercept 1 public 1 gpa 2.9989 / ilink category='0';
estimate "Pr prob apply=0 at pared=1" intercept 1 pared 1 public 1 gpa 2.9989 / ilink category='0';
run;
```

*****SOME OUTPUT OMITTED AND LAYOUT MODIFIED*****

Label APPLY Estimate Standard Error z Value Pr > |z| Mean Standard Error of Mean
Pr prob apply=0 at pared=0 0 0.4153 0.2733 1.52 0.1286 0.6024 0.06546
Pr prob apply=0 at pared=1 0 -0.6325 0.3451 -1.83 0.0668 0.3469 0.07818

Things to consider

See also

References