Ordinal Logistic Regression

Department of Public Health & Health Policy

November 4 2022

Content

- Introduction
- Formulation
- Assumptions
- Types of Procedures
- Demonstration in SPSS
- Results

Introduction

- **Ordinal logistic regression** (often just called 'ordinal regression')
  - used to predict an ordered(ranked) dependent variable with one or more independent variables.
- Use interactions between independent variables to predict the dependent variable
- Determine which of your independent variables (if any) have a statistically significant effect on your dependent variable.
- Determine how well your ordinal regression model predicts the dependent variable.

Formulation

- If dependent variable has k categories, it is cumulatively split into k-1 variables binomial(dichotomous) variables
- Binomial variables j separates the cases in the first j categories of the dependent variable (1, 2, 3, …, j) from the other (j+1, j+2,…,k)
- For each binomial variable j, the binomial regression is as follows

$$ln\frac{p_j}{1-p_j} = b_{0j} + b_{1j}x_1 + b_{2j}x_2 + \cdots + b_{kj}x_k$$

- Pj is the probability of success for the variable j, while the b's are the regression coefficients

Assumptions

1. **Dependent variable** should be measured at the **ordinal** level
   - E.g. Likert items (e.g., a 7-point scale from "strongly agree" through to "strongly disagree"),
   - a 3-point scale (customer liked a product, ranging from "Not very much", to "It is OK", to "Yes, a lot")

2. One or more **independent variables** are **continuous**, **ordinal** or **categorical** (including **dichotomous variables**).
   - However, ordinal independent variables must be treated as being either continuous or categorical.

3. There is **no multicollinearity**

4. You have **proportional odds**
   - The coefficients b1j, b2j upto bkj are equal for all the (k-1) logistic regressions

(2015, Osborne)

Testing assumptions #3 and #4

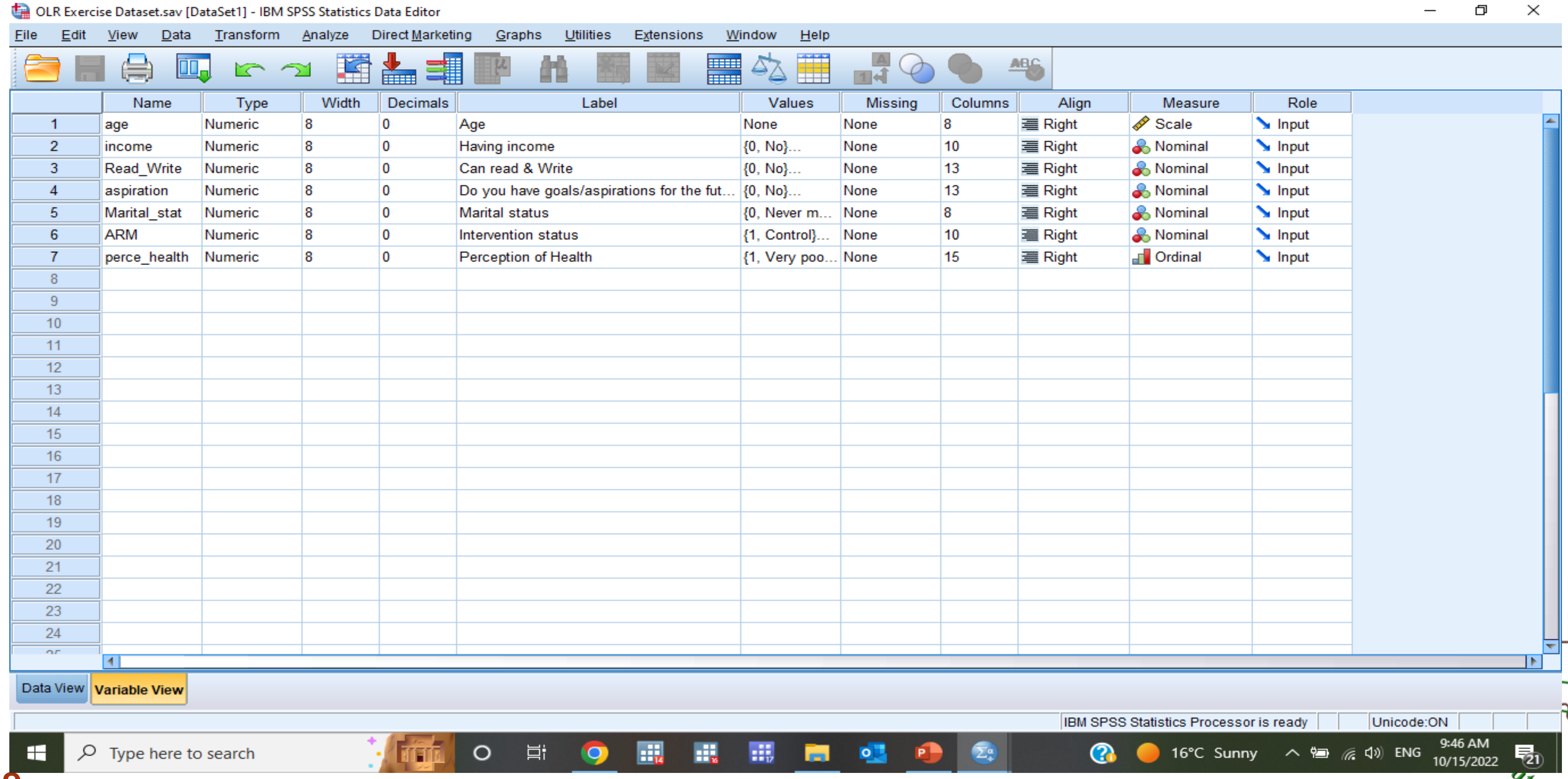
- Assumption #3. Multi-collinearity
  - Independent variables must be continuous or categorical, but not ordinal
  - Check VIF
- Assumption #4. Proportional odds
  - Test of parallel lines
    - A non-significant test result suggest the assumption of proportional odds is met, meaning the effects of the independent variables on the cumulative probability of falling into a higher category doesn't vary across categories on the dependent variable
    - P value should be $p>0.05$

Types of Procedures

PLUM (Polytomous Universal Model)	GENLIN (Generalized Linear Model)
Available on Basic SPSS modules	Available on advanced SPSS modules
Doesn’t automatically compute antilogarithms of the regression coefficient	More easy and quicker
	But, it doesn’t allow the test of proportional odds, doesn't compute Pseudo R-square, and it doesn’t save the predicted probabilities in a convenient way

Demonstration

Result



OLR Exercise Dataset.sav [DataSet1] - IBM SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	age	Numeric	8	0	Age	None	None	8	Right	Scale	Input
2	income	Numeric	8	0	Having income	{0, No}...	None	10	Right	Nominal	Input
3	Read_Write	Numeric	8	0	Can read & Write	{0, No}...	None	13	Right	Nominal	Input
4	aspiration	Numeric	8	0	Do you have goals/aspirations for the fut...	{0, No}...	None	13	Right	Nominal	Input
5	Marital_stat	Numeric	8	0	Marital status	{0, Never m...	None	8	Right	Nominal	Input
6	ARM	Numeric	8	0	Intervention status	{1, Control}...	None	10	Right	Nominal	Input
7	perce_health	Numeric	8	0	Perception of Health	{1, Very poo...	None	15	Right	Ordinal	Input

Model-fitness test

*Output4 [Document4] - IBM SPSS Statistics Viewer

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	937.420			
Final	902.269	35.152	6	.000

Link function: Logit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	387.798	398	.633
Deviance	377.658	398	.761

Link function: Logit.

Pseudo R-Square

Cox and Snell	.011
Nagelkerke	.012
McFadden	.005

Link function: Logit.

Parallel odds

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	902.269			
General	882.073	20.196	18	.322

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

Multi-collinearity test

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Intervention status, Marital status, Do you have goals/aspirations for the future?, Can read & Write, Having income, Age[b]	.	Enter

a. Dependent Variable: Perception of Health

b. All requested variables entered.

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Age	.872	1.147
	Having income	.938	1.066
	Can read & Write	.953	1.049
	Do you have goals/aspirations for the future?	.980	1.020
	Marital status	.858	1.166
	Intervention status	.988	1.012

a. Dependent Variable: Perception of Health

OLR result

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Threshold	[Perception of Health=1]	-3.479	.4631	-4.387	-2.571	56.448	1	.000	.031	.012	.076
	[Perception of Health=2]	-2.075	.4458	-2.949	-1.202	21.671	1	.000	.126	.052	.301
	[Perception of Health=3]	-.703	.4419	-1.569	.163	2.534	1	.111	.495	.208	1.177
	[Perception of Health=4]	.540	.4417	-.326	1.406	1.496	1	.221	1.716	.722	4.079
[Intervention status=2]		.359	.0713	.219	.498	25.305	1	.000	1.431	1.245	1.646
[Intervention status=1]		0[a]	.	.	.	.	.	.	1	.	.
Age		.014	.0276	-.040	.068	.246	1	.620	1.014	.960	1.070
[Can read & Write=1]		.207	.0787	.053	.362	6.936	1	.008	1.230	1.054	1.436
[Can read & Write=0]		0[a]	.	.	.	.	.	.	1	.	.
[Marital status=1]		.107	.1460	-.179	.393	.534	1	.465	1.113	.836	1.481
[Marital status=0]		0[a]	.	.	.	.	.	.	1	.	.
[Having income=1]		.057	.1125	-.164	.277	.252	1	.616	1.058	.849	1.319
[Having income=0]		0[a]	.	.	.	.	.	.	1	.	.
[Do you have goals/aspirations for the future?=1]		.173	.1801	-.180	.526	.923	1	.337	1.189	.835	1.692
[Do you have goals/aspirations for the future?=0]		0[a]	.	.	.	.	.	.	1	.	.
(Scale)		1[b]									

Dependent Variable: Perception of Health
Model: (Threshold), Intervention status, Age, Can read & Write, Marital status, Having income, Do you have goals/aspirations for the future?

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Interpretation

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper	Hypothesis Test Wald Chi-Square	Hypothesis Test df	Hypothesis Test Sig.	Exp(B)	95% Wald Confidence Interval for Exp(B) Lower	95% Wald Confidence Interval for Exp(B) Upper
[Intervention status=2]	.359	.0713	.219	.498	25.305	1	.000	1.431	1.245	1.646
[Intervention status=1]	0[a]	.	.	.	.	.	.	1	.	.
[Can read & Write=1]	.207	.0787	.053	.362	6.936	1	.008	1.230	1.054	1.436
[Can read & Write=0]	0[a]	.	.	.	.	.	.	1	.	.

The odds of girls in the intervention area being in the higher group of health condition status was **1.43(95%CI, 1.24 to 1.65**) times that of girls in the control area.

- The Probability that a girl from the intervention area to perceive that she has “Very good” health condition is 58.8%.

The odds of girls who can read and write being in the higher group of health condition status was **1.23(95%CI, 1.05 to 1.43)** times that of girls who can’t read and write..

- The Probability that a girl who can read and write to perceive that she has “Very good” health condition is 55.2%

References

- Ordinal Logistic Regression models and Statistical Software: What You Need to Know. Cornel University, 2020
- https://stats.oarc.ucla.edu/spss/dae/ordinal-logistic-regression/
- Fagerland M. How to test for goodness of fit in ordinal logistic regression models, 2017