Vanderbilt University
Leadership, Policy and Organizations
Class Number 9553
Spring 2017

Models for Categorical Outcomes

Ordered Outcomes

We many times work with data that is categorical (mutually exclusive and exhaustive categories), but ordered— each level implies going up or down from the previous level. What makes this ordinal data is that the steps do not imply the same distance on a given scale. For instance, consider an outcome like educational attainment, coded as less than high school, high school, some college, bachelor's degree, advanced degree. While these are clearly ordered—each step implies going up the educational ladder— the distance between steps is not as simple as the number of years of education. For this situation we can use the ordered probit model.

Econometric Model

The random utiliity form of the ordered probit model is as follows:

$$\boldsymbol{y}_i^* = \boldsymbol{x\beta} + e_i, \qquad e_i \sim N(0,1)\, i = 1 \dots N$$

y_i, the observed variable takes on values 0 through m:

$$y_i = j \iff \tau_{j-1} < y_i^* \leq \tau_j$$

For the ordinal probit model, the assumption is that e_i is distributed normally with mean 0, and typically standard deviation 1.

We then will have defined y^* as being within a normal distribution, with the probability space from 0 to 1 divided by a set of cut point τ_k

The probability that y=1 is given by:

$$Pr(y_i = 1|\boldsymbol{x}_i) = Pr(\tau_0 \leq y_i^* < \tau_1|\boldsymbol{x})$$

Substituting from above and evaluating the cdf of the normal distribution (ϕ) to calculate probabilities within each space gives use the following.

To find the probability that $y = m$, we utilize the following equation

$$P(y_i = m) = \phi(\tau_m - \boldsymbol{x}_i\boldsymbol{\beta}) - \phi(\tau_{m-1} - \boldsymbol{x}_i\boldsymbol{\beta})$$

This is actually a generalization of the probit model for binary response– in the binary case, m can be 0 or 1, while in the ordered case, m can take on any number of values.

This model is not identified without an additional assumption. We can either set the intercept in the linear model to be constant, usually at 0, ($\beta_0 = 0$) or we can assume that τ_1=0. By default, Stata holds ($\beta_0 = 0$).

Reporting Results

These models can be difficult to interpret for the same reason the models for binary data can be difficult: the interpretation is on the logit scale or the normal scale, and parameter estimates cannot be interpreted one at a time in terms of probability. I present two options below.

Consider the following set of results from an ordinal probit regression of college plans (Don't know=1, Two year=2, Four year=3) on a set of student characteristics.

Table 1: Results of Ordinal Probit Estimates, Dependent Variable= College Plans (None, 2yr, 4yr)

	(1)	(2)	(3)
	order_plan	order_plan	order_plan
order_plan			
female	0.29	0.29	0.31
	(0.02)	(0.02)	(0.02)
bynels2m	2.52	2.70	2.37
	(0.13)	(0.13)	(0.13)
bynels2r	2.66	2.94	2.38
	(0.18)	(0.19)	(0.19)
amind		0.04	0.13
		(0.12)	(0.12)
asian		0.35	0.44
		(0.04)	(0.04)
black		0.48	0.55
		(0.04)	(0.04)
hispanic		0.05	0.19
		(0.03)	(0.03)
multiracial		-0.00	0.02
		(0.05)	(0.05)
byses1			0.38
			(0.02)
cut1			
_cons	0.46	0.70	0.42
	(0.04)	(0.05)	(0.05)
cut2			
_cons	1.68	1.94	1.68
	(0.04)	(0.05)	(0.05)
N	13055	13055	13055
AIC	19165.26	18947.23	18493.51

Standard errors in parentheses

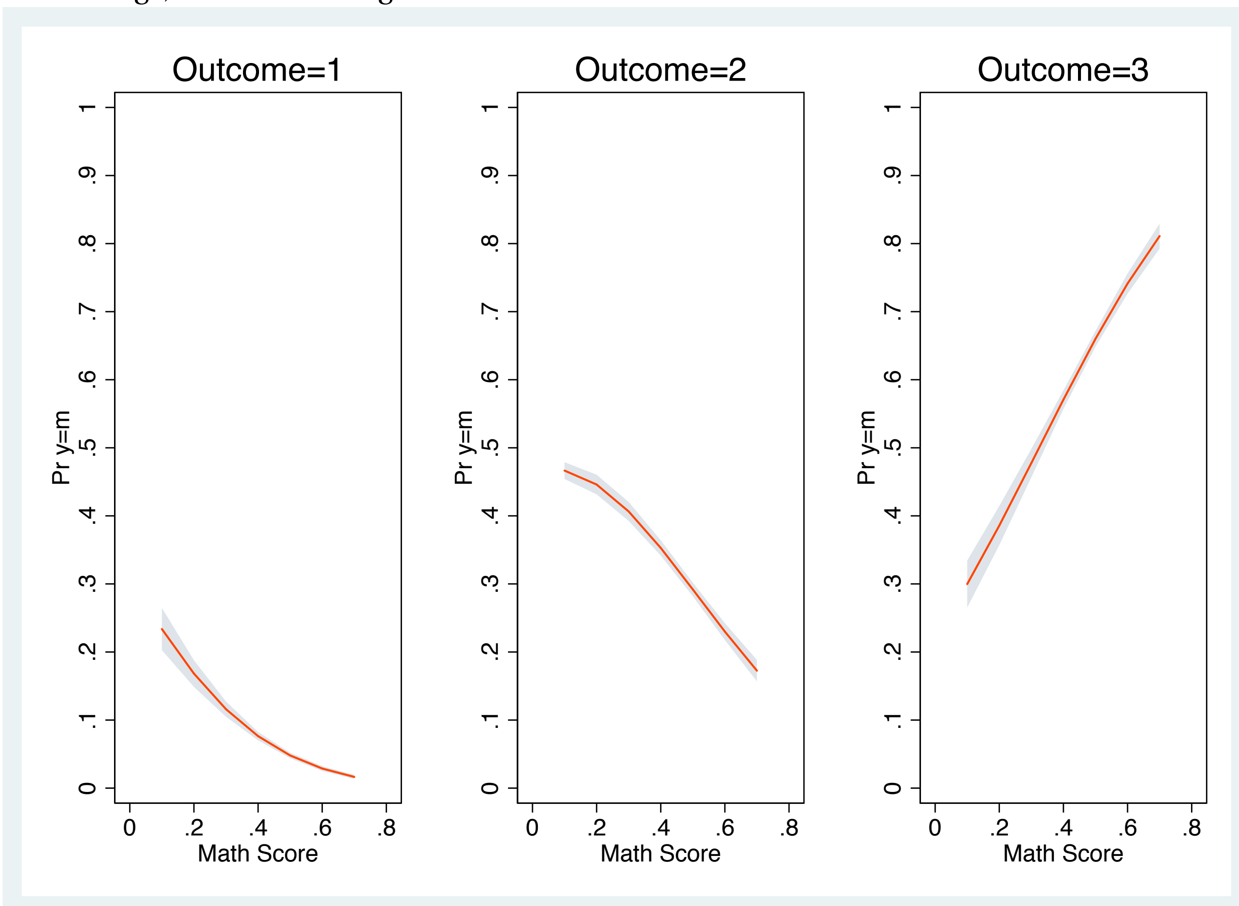
Table 2: Marginal Effects for Variables, Dependent Variable= College Plans (None, 2yr, 4yr) Outcomes in Columns

	(1)	(2)	(3)
bynels2m	-0.28	-0.44	0.72
	[-0.31,-0.25]	[-0.49,-0.40]	[0.65,0.80]
bynels2r	-0.28	-0.45	0.72
	[-0.32,-0.23]	[-0.51,-0.38]	[0.61,0.83]
amind	-0.02	-0.02	0.04
	[-0.04,0.01]	[-0.07,0.02]	[-0.03,0.11]
asian	-0.05	-0.08	0.13
	[-0.06,-0.04]	[-0.10,-0.07]	[0.11,0.16]
black	-0.06	-0.10	0.17
	[-0.07,-0.06]	[-0.12,-0.09]	[0.15,0.19]
hispanic	-0.02	-0.04	0.06
	[-0.03,-0.01]	[-0.05,-0.02]	[0.04,0.08]
multiracial	-0.00	-0.00	0.01
	[-0.02,0.01]	[-0.02,0.02]	[-0.03,0.04]
byses1	-0.04	-0.07	0.12
	[-0.05,-0.04]	[-0.08,-0.06]	[0.11,0.13]
N	13055	13055	13055

95% confidence intervals in brackets

Without additional interpretation, this doesn't make sense. One way to interpret results is to use marginal effects, calculated for each outcome, holding all other variables at their means.

Figure 1: Predicted Impact of Math Test Scores on College Plans, Outcome= Don't Know, Two Year College, Four Year College



As always, I reccomend plotting the data. The figure below shows the predicted impact of test scores on the probability of each outcome, holding all other variables constant at their minimum for binary variables and at their mean for continuous variables.

Quick Exercise

Conduct a regression with institution of first attendance as an ordered outcome and report on the results.

Unordered Outcomes

When dealing with unordered categorical data, we no longer have the information that the outcomes share some agreed order. Instead, we have categorical information. A classic example is the type of college a student attends: public four year, private four year, public two year. The choices are not ordered in any way that makes sense. The model here is more complicated. We need to model tradeoffs between each of the choices. The model here takes the form:

$$\begin{aligned} Pr(Y_i = 1) &= Pr(\mu_{i1} > \mu_{i2}) \\ &= Pr(\mu_{i1} + e_{i1} > \mu_{i2} + e_{i2}) \\ &= Pr(e_{i1} - e_{i2} > \mu_{i2} - \mu_{i1}) \end{aligned}$$

In effect, we are comparing each of the choices with one another, with utility set, as always, to = $\boldsymbol{x\beta}$

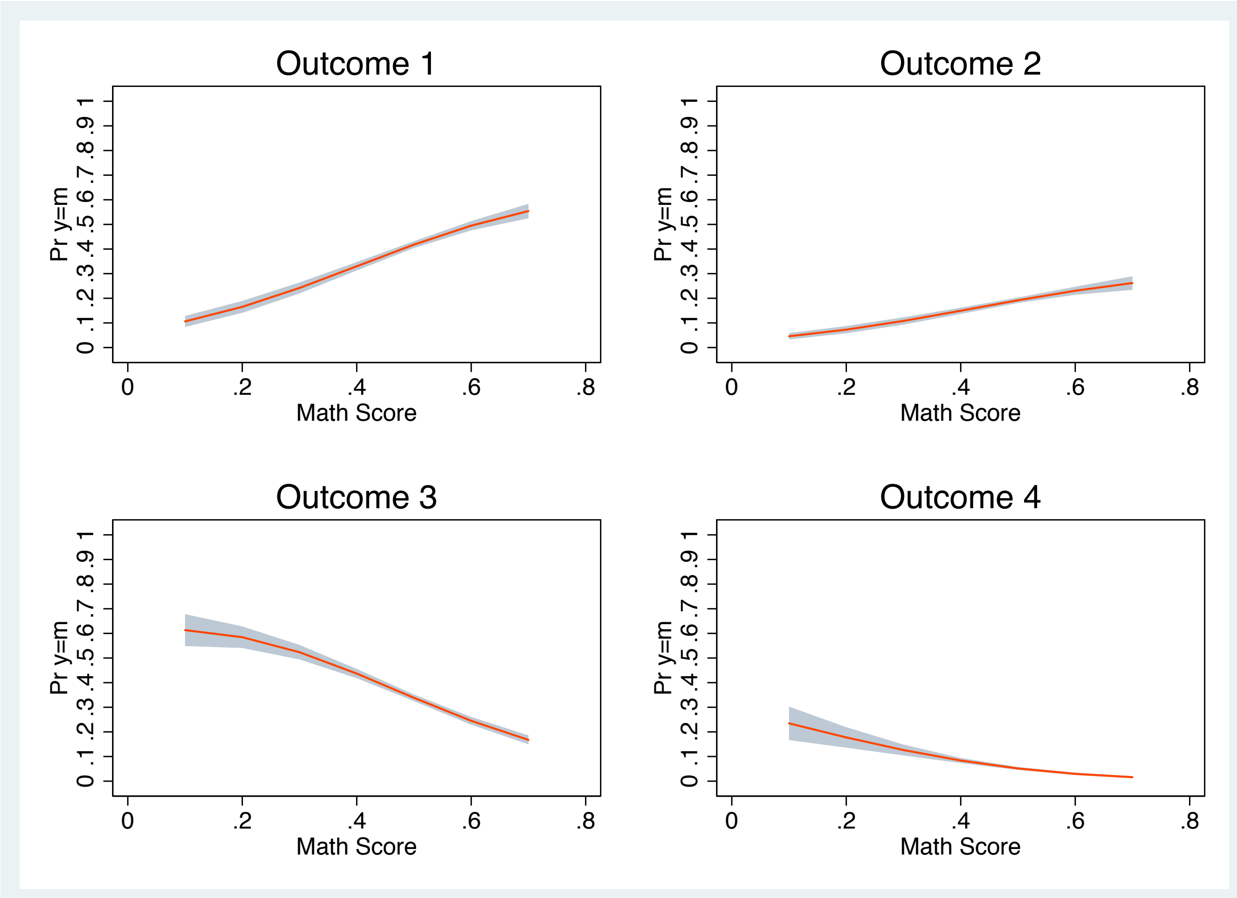
$$Pr(y = m) = Pr(\boldsymbol{x\beta}_m > \boldsymbol{x\beta}_j \forall j \neq m)$$

This means that all coefficients will be *relative*. Below are the coefficients for a model of first instiitution attended, with public two year, other and private four year compared to public four year.

Table 3: Results of Multinomial Logit model, outcome= First College Attended

	(1)	(2)	(3)
	first_inst	first_inst	first_inst
Public_4_Year			
female	0.00	0.00	0.00
	(.)	(.)	(.)
bynels2m	0.00	0.00	0.00
	(.)	(.)	(.)
bynels2r	0.00	0.00	0.00
	(.)	(.)	(.)
amind		0.00	0.00
		(.)	(.)
asian		0.00	0.00
		(.)	(.)
black		0.00	0.00
		(.)	(.)
hispanic		0.00	0.00
		(.)	(.)
multiracial		0.00	0.00
		(.)	(.)
byses1			0.00
			(.)
_cons	0.00	0.00	0.00
	(.)	(.)	(.)
Private_4_Year			
female	0.04	0.04	0.06
	(0.06)	(0.06)	(0.06)
bynels2m	0.31	0.50	0.15
	(0.33)	(0.35)	(0.35)
bynels2r	2.72***	2.39***	1.88***
	(0.47)	(0.47)	(0.48)
amind		0.34	0.44
		(0.41)	(0.41)
asian		-0.39***	-0.34***
		(0.09)	(0.09)
black		-0.05	0.02
		(0.10)	(0.10)
hispanic		-0.16	-0.07
		(0.11)	(0.11)
multiracial		0.13	0.15
		(0.13)	(0.13)
byses1			0.36***
			(0.04)
_cons	-1.81***	-1.75***	-1.57***
	(0.16)	(0.17)	(0.17)
Public_2_Year			
female	-0.17***	-0.18***	-0.22***
	(0.05)	(0.05)	(0.05)
bynels2m	-4.92***	-5.22***	-4.94***
	(0.29)	(0.30)	(0.31)
bynels2r	-4.38***	-4.66***	-4.02***
	(0.40)	(0.41)	(0.42)
amind		0.10	-0.04
		(0.35)	(0.35)
asian		-0.22**	-0.35***
		(0.08)	(0.09)
black		-0.70***	-0.83***
		(0.09)	(0.09)
hispanic		0.21*	0.04
		(0.08)	(0.08)
multiracial		0.01	-0.01
		(0.13)	(0.13)
byses1			-0.46***
			(0.04)
_cons	3.64***	3.96***	3.76***
	(0.12)	(0.14)	(0.14)
Other			
female	-0.29**	-0.29**	-0.36***
	(0.09)	(0.09)	(0.09)
bynels2m	-7.85***	-7.63***	-7.25***
	(0.51)	(0.53)	(0.53)
bynels2r	-3.60***	-3.97***	-3.04***
	(0.71)	(0.72)	(0.73)
amind		0.49	0.25
		(0.48)	(0.49)
asian		-0.65***	-0.88***
		(0.18)	(0.19)

Figure 2: Predicted Impact of Math Test Scores on First Institution Attended Outcome=Public 4 Year, Private 4 Year, Public 2 Year, Other



Quick Exercise

Conduct a regression with a three-part definition of first institution and report on the results.

As with ordered outcome, the key to interpreting results for a multinomial model is to predict probabilites and compare those predicted probabilities. As we'll see in the case of first institution choice, the same prospective students will have very different predicted probabilities of attendance at different types of institutions.

Measuring model fit for models for ordered outcomes is almost always done using the log likelihood. One can use the likelihood ratio test, AIC or BIC to summarize model fit. There are ways to implement the AUC for a multiclass outcome, but, it's complicated.