Ordinal and Multinomial Logistic Regression

Andy Grogan-Kaylor

19 Oct 2020

Meta-Background

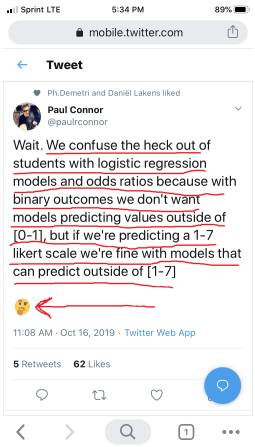

Figure 1: Tweet About Ordinal Models

Key Concepts and Commands

- Implementations differ; formulas are our friends
- Extensions to logistic model: ordinal and multinomial logit

$$F(y) = \beta_0 + \beta x_1 + \beta x_2 + ...$$

- Ordinal model

$y(1, 2, 3, \text{etc.}) = \beta_0 + \beta x_1 + \beta x_2 + ...$

- Multinomial model

$y(2 \text{ vs. } 1) = \beta_0 + \beta x_1 + \beta x_2 + ...$
$y(3 \text{ vs. } 1) = \beta_0 + \beta x_1 + \beta x_2 + ...$

- Think about OR's, predicted probabilities, non-linearity
- Different models for *different types* of ordinal variables

Get The Data (General Social Survey)

```
. clear all

. set maxvar 10000 // increase number of allowable variables

. use "/Users/agrogan/Box Sync/DATA WAREHOUSE/General Social Survey/GSS7218_R1.DTA", clear

. keep polviews sex maeduc paeduc age degree coninc

. save GSSsmall.dta, replace
file GSSsmall.dta saved

. describe // describe the data
Contains data from GSSsmall.dta
  obs:        64,814
 vars:             7                          19 Oct 2020 08:01
 size:       907,396
```

variable name	storage type	display format	value label	variable label
age	byte	%8.0g	AGE	age of respondent
paeduc	byte	%8.0g	LABK	highest year school completed, father
maeduc	byte	%8.0g	LABK	highest year school completed, mother
degree	byte	%8.0g	LABL	r´s highest degree
sex	byte	%8.0g	SEX	respondents sex
polviews	byte	%8.0g	POLVIEWS	think of self as liberal or conservative
coninc	double	%12.0g	LABIH	family income in constant dollars

```
Sorted by:
```

Thinking About Your Data and Data Wrangling

It is always good to think about your data and what the values of different variables represent. In Stata, however, there is very little additional data wrangling to prepare the data. In R, there is considerable data wrangling since we have to employ special commands just to get *variable* and *value* labels, and to ensure that *numeric dependent* variables are recoded as *factors*. In Stata there are no such issues!!!

Descriptive Statistics

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	64,586	46.09936	17.5347	18	89
paeduc	45,837	10.71026	4.342689	0	20
maeduc	53,870	10.85365	3.768792	0	20
degree	64,641	1.35858	1.175289	0	4
sex	64,814	1.558521	.4965673	1	2
polviews	55,328	4.100528	1.382474	1	7
coninc	58,294	45028.17	36791	350.5	180386

```
. tabulate polviews
```

think of self as liberal or conservative	Freq.	Percent	Cum.
extremely liberal	1,682	3.04	3.04

liberal	6,514	11.77	14.81
slightly liberal	7,010	12.67	27.48
moderate	21,370	38.62	66.11
slghtly conservative	8,690	15.71	81.81
conservative	8,230	14.87	96.69
extrmly conservative	1,832	3.31	100.00
Total	55,328	100.00	

The Ordinal Model (*k categories*)[1]

$$p_{ij} = Pr\Big(k_{i-1} < x_j\beta <= k_i\Big)$$

$$\ln\left(\frac{p(y \leq k)}{p(y > k)}\right) = \beta_0 + \beta_1 x_1 + ...$$

Ordinal Regression

```
. ologit polviews sex age degree coninc
Iteration 0:   log likelihood = -83895.058
Iteration 1:   log likelihood = -83369.429
Iteration 2:   log likelihood = -83368.485
Iteration 3:   log likelihood = -83368.485
Ordered logistic regression                     Number of obs     =     50,049
                                                LR chi2(4)        =    1053.15
                                                Prob > chi2       =     0.0000
Log likelihood = -83368.485                     Pseudo R2         =     0.0063
```

polviews	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex	-.129234	.0162348	-7.96	0.000	-.1610536	-.0974144
age	.0116653	.0004737	24.63	0.000	.0107369	.0125937
degree	-.1062661	.0076242	-13.94	0.000	-.1212093	-.091323
coninc	3.99e-06	2.42e-07	16.52	0.000	3.52e-06	4.46e-06
/cut1	-3.116098	.0440989			-3.202531	-3.029666
/cut2	-1.389623	.0379027			-1.463911	-1.315335
/cut3	-.5941761	.0372164			-.6671188	-.5212333
/cut4	1.050951	.037438			.9775742	1.124329
/cut5	1.916652	.03824			1.841703	1.991601
/cut6	3.826484	.0447146			3.738845	3.914123

Many commands for regression of categorical dependent variables in R *do not provide p values*, and an extra step has to be taken to get p values. This is *not* a problem in Stata!

Exponentiating Coefficients: e^{β}

```
. ologit polviews sex age degree coninc, or
Iteration 0:   log likelihood = -83895.058
Iteration 1:   log likelihood = -83369.429
Iteration 2:   log likelihood = -83368.485
Iteration 3:   log likelihood = -83368.485
Ordered logistic regression                     Number of obs     =     50,049
```

[1]Per Stata documentation.

```
                                                LR chi2(4)        =    1053.15
                                                Prob > chi2       =     0.0000
Log likelihood = -83368.485                     Pseudo R2         =     0.0063
```

polviews	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex	.8787683	.0142666	-7.96	0.000	.8512464	.90718
age	1.011734	.0004792	24.63	0.000	1.010795	1.012673
degree	.8991853	.0068555	-13.94	0.000	.8858486	.9127228
coninc	1.000004	2.42e-07	16.52	0.000	1.000004	1.000004
/cut1	-3.116098	.0440989			-3.202531	-3.029666
/cut2	-1.389623	.0379027			-1.463911	-1.315335
/cut3	-.5941761	.0372164			-.6671188	-.5212333
/cut4	1.050951	.037438			.9775742	1.124329
/cut5	1.916652	.03824			1.841703	1.991601
/cut6	3.826484	.0447146			3.738845	3.914123

```
Note: Estimates are transformed only in the first equation.
```

The Proportional Odds Assumption And The Brant Test

```
. brant
Brant test of parallel regression assumption
```

	chi2	p>chi2	df
All	1456.59	0.000	20
sex	108.03	0.000	5
age	120.63	0.000	5
degree	835.26	0.000	5
coninc	67.78	0.000	5

```
A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

The Multinomial Model

$$\ln\left(\frac{P(y=y_2)}{P(y=y_1)}\right) = \ln\left(\frac{P(y=\text{something else})}{P(y=\text{something})}\right)$$

$$= \beta_0 + \beta_1 x_1 + ...$$

$$\ln\left(\frac{P(y=y_3)}{P(y=y_1)}\right) = \ln\left(\frac{P(y=\text{something else altogether})}{P(y=\text{something})}\right)$$

$$= \beta_0 + \beta_1 x_1 + ...$$

Estimation

```
. mlogit polviews i.sex age degree coninc
Iteration 0:   log likelihood = -83895.058
Iteration 1:   log likelihood = -82700.548
Iteration 2:   log likelihood = -82694.595
Iteration 3:   log likelihood = -82694.594
Multinomial logistic regression                 Number of obs     =     50,049
```

```
                                                LR chi2(24)       =    2400.93
                                                Prob > chi2       =     0.0000
Log likelihood = -82694.594                     Pseudo R2         =     0.0143
```

polviews	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
extremely_liberal						
sex						
female	-.2153043	.0534275	-4.03	0.000	-.3200202	-.1105883
age	-.0051601	.0015774	-3.27	0.001	-.0082517	-.0020685
degree	.3607061	.0234865	15.36	0.000	.3146735	.4067387
coninc	-6.68e-06	8.90e-07	-7.51	0.000	-8.43e-06	-4.94e-06
_cons	-2.40105	.0904486	-26.55	0.000	-2.578326	-2.223774
liberal						
sex						
female	-.0770042	.0302144	-2.55	0.011	-.1362233	-.0177851
age	-.0077271	.0009041	-8.55	0.000	-.0094991	-.0059551
degree	.3615385	.0134905	26.80	0.000	.3350977	.3879794
coninc	-2.36e-06	4.59e-07	-5.14	0.000	-3.26e-06	-1.46e-06
_cons	-1.195919	.0513843	-23.27	0.000	-1.29663	-1.095207
slightly_liberal						
sex						
female	-.1016619	.0292053	-3.48	0.000	-.1589032	-.0444206
age	-.0099768	.0008799	-11.34	0.000	-.0117014	-.0082521
degree	.2358701	.0134562	17.53	0.000	.2094964	.2622438
coninc	-1.94e-07	4.37e-07	-0.44	0.658	-1.05e-06	6.63e-07
_cons	-.90455	.0494119	-18.31	0.000	-1.001396	-.8077044
moderate	(base outcome)					
slghtly_conservative						
sex						
female	-.2630355	.0270206	-9.73	0.000	-.315995	-.210076
age	.0012542	.0007943	1.58	0.114	-.0003026	.002811
degree	.1963805	.012493	15.72	0.000	.1718947	.2208663
coninc	3.39e-06	3.86e-07	8.79	0.000	2.63e-06	4.15e-06
_cons	-1.221032	.0467118	-26.14	0.000	-1.312585	-1.129479
conservative						
sex						
female	-.2625249	.0278997	-9.41	0.000	-.3172073	-.2078426
age	.0128524	.000801	16.05	0.000	.0112825	.0144224
degree	.152561	.0129671	11.77	0.000	.127146	.177976
coninc	3.87e-06	3.97e-07	9.75	0.000	3.09e-06	4.65e-06
_cons	-1.813802	.0496044	-36.57	0.000	-1.911025	-1.716579
extrmly_conservative						
sex						
female	-.3790287	.0530006	-7.15	0.000	-.482908	-.2751493
age	.0150308	.0014834	10.13	0.000	.0121235	.0179381
degree	.004062	.0262081	0.15	0.877	-.0473049	.055429
coninc	3.35e-07	8.19e-07	0.41	0.682	-1.27e-06	1.94e-06
_cons	-3.040997	.0945989	-32.15	0.000	-3.226407	-2.855587

Exponentiating Coefficients

```
. mlogit, rr
Multinomial logistic regression                 Number of obs     =     50,049
                                                LR chi2(24)       =    2400.93
                                                Prob > chi2       =     0.0000
Log likelihood = -82694.594                     Pseudo R2         =     0.0143
```

polviews	RRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]

extremely_liberal						
sex						
female	.8062961	.0430784	-4.03	0.000	.7261343	.8953073
age	.9948532	.0015693	-3.27	0.001	.9917823	.9979336
degree	1.434342	.0336876	15.36	0.000	1.369812	1.501912
coninc	.9999933	8.90e-07	-7.51	0.000	.9999916	.9999951
_cons	.0906228	.0081967	-26.55	0.000	.075901	.1082
liberal						
sex						
female	.925886	.0279751	-2.55	0.011	.8726477	.9823721
age	.9923027	.0008971	-8.55	0.000	.9905458	.9940626
degree	1.435536	.0193661	26.80	0.000	1.398077	1.473999
coninc	.9999976	4.59e-07	-5.14	0.000	.9999967	.9999985
_cons	.3024259	.01554	-23.27	0.000	.2734517	.3344702
slightly_liberal						
sex						
female	.9033349	.0263822	-3.48	0.000	.8530789	.9565515
age	.9900729	.0008712	-11.34	0.000	.9883668	.9917818
degree	1.26601	.0170357	17.53	0.000	1.233057	1.299843
coninc	.9999998	4.37e-07	-0.44	0.658	.9999989	1.000001
_cons	.404724	.0199982	-18.31	0.000	.3673664	.4458805
moderate	(base outcome)					
slghtly_conservative						
sex						
female	.7687146	.0207712	-9.73	0.000	.7290631	.8105226
age	1.001255	.0007953	1.58	0.114	.9996975	1.002815
degree	1.21699	.0152038	15.72	0.000	1.187553	1.247157
coninc	1.000003	3.86e-07	8.79	0.000	1.000003	1.000004
_cons	.2949256	.0137765	-26.14	0.000	.2691234	.3232017
conservative						
sex						
female	.7691072	.0214578	-9.41	0.000	.7281798	.8123349
age	1.012935	.0008114	16.05	0.000	1.011346	1.014527
degree	1.164814	.0151042	11.77	0.000	1.135583	1.194797
coninc	1.000004	3.97e-07	9.75	0.000	1.000003	1.000005
_cons	.1630332	.0080872	-36.57	0.000	.1479287	.1796798
extrmly_conservative						
sex						
female	.684526	.0362803	-7.15	0.000	.6169866	.7594587
age	1.015144	.0015058	10.13	0.000	1.012197	1.0181
degree	1.00407	.0263148	0.15	0.877	.9537966	1.056994
coninc	1	8.19e-07	0.41	0.682	.9999987	1.000002
_cons	.0477872	.0045206	-32.15	0.000	.0396999	.0575221

Note: _cons estimates baseline relative risk for each outcome.

Predicted Probabilities

```
. margins sex, predict(outcome(1)) // predicted probabilities by sex; y = 1
Predictive margins                              Number of obs     =     50,049
Model VCE    : OIM
Expression   : Pr(polviews==extremely_liberal), predict(outcome(1))
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex						
male	.0325114	.001187	27.39	0.000	.0301849	.0348378
female	.0295928	.0010205	29.00	0.000	.0275927	.031593