Queensland University of Technology
Transport Data Analysis and Modeling Methodologies

Lab Session #15a
(Ordered Discrete Data – With a Multivariate Binary Probit Model)
Based on Example 14.1

A survey of 250 commuters was in the Seattle metropolitan area (this sample is reduced from the 322 given in the book due to the elimination of some missing data). The survey's intent was to gather information on commuters' opinions of high-occupancy vehicle (HOV) lanes (lanes that are restricted for use by vehicles with 2 or more occupants). The variables available from this survey are given on the attached table.

Among the questions asked, commuters were asked whether they agreed with the following statements:

1. "HOV lanes save all commuters time" (variable number x27 in the data table)
2. "Existing HOV lanes are being adequately used." (variable number x28 in the data table)
3. "HOV lanes should be open to all vehicles, regardless of vehicle occupancy level" (variable number x29 in the table).
4. "Converting some regular lanes to HOV lanes is a good idea" (variable number x30 in the data table).
5. "Converting some regular lanes to HOV lanes is a good idea only if it is done before traffic congestion becomes serious" (variable number x31 in the data table).

The question provided ordered responses of; strongly disagree, disagree, neutral, agree, agree strongly. But suppose we are interested in whether respondents disagree or not, so that we have just two outcomes: disagree (disagree or strongly disagree) or do not disagree (neutral, agree, agree strongly). With this, note that these five questions are obviously interrelated. To understand the factors determining these five commuter opinions, a multivariate binary probit model of these survey questions is appropriate (with the original data recoded to disagree/do-not-disagree as described above).

Your task is to estimate a multivariate model of the five response variables mentioned above.

1. The results of your best model specification.

2. A discussion of the logical process that led you to the selection of your final specification (discuss the theory behind the inclusion of your selected variables). Include t-statistics and justify the sign of your variables.

Variables available for your specification are (in file Ex14-1.txt):

Variable Number	Explanation
x1	Usual mode of travel: 0 if drive alone, 1 if two person carpool, 2 if three or more person carpool, 3 if vanpool, 4 if bus, 5 if bicycle or walk, 6 if motorcycle, 7 if other
x2	Have used HOV lanes: 1 if yes, 0 if no
x3	If used HOV lanes, what mode is most often used: 0 in a bus, 1 in two person carpool, 2 in three or more person carpool, 3 in vanpool, 4 alone in vehicle, 5 on motorcycle
x4	Sometimes eligible for HOV lane use but do not use: 1 if yes, 0 if no
x5	Reason for not using HOV lanes when eligible: 0 if slower than regular lanes, 1 if too much trouble to change lanes, 2 if HOV lanes are not safe, 3 if traffic moves fast enough, 4 if forget to use HOV lanes, 5 if other
x6	Usual mode of travel one year ago: 0 if drive alone, 1 if two person carpool, 2 if three or more person carpool, 3 if vanpool, 4 if bus, 5 if bicycle or walk, 6 if motorcycle, 7 if other
x7	Commuted to work in Seattle a year ago: 1 if yes, 0 if no
x8	Have flexible work start times: 1 if yes, 0 if no
x9	Changed departure times to work in the last year: 1 if yes, 0 if no
x10	On average, number of minutes leaving earlier for work relative to last year
x11	On average, number of minutes leaving later for work relative to last year
x12	If changed departure times to work in the last year, reason why: 0 if change in travel mode, 1 if increasing traffic congestion, 2 if change in work start time, 3 if presence of HOV lanes, 4 if change in residence, 5 if change in lifestyle, 6 if other
x13	Changed route to work in the last year: 1 if yes, 0 if no
x14	If changed route to work in the last year, reason why: 0 if change in travel mode, 1 if increasing traffic congestion, 2 if change in work start time, 3 if presence of HOV lanes, 4 if change in residence, 5 if change in lifestyle, 6 if other
x15	Usually commute to or from work on Interstate 90: 1 if yes, 0 if no

x16	Usually commuted to or from work on Interstate 90 last year: 1 if yes, 0 if no
x17	On your past five commutes to work, how often have you used HOV lanes
x18	On your past five commutes to work, how often did you drive alone
x19	On your past five commutes to work, how often did you carpool with one other person
x20	On your past five commutes to work, how often did you carpool with two or more people
x21	On your past five commutes to work, how often did you take a vanpool
x22	On your past five commutes to work, how often did you take a bus
x23	On your past five commutes to work, how often did you bicycle or walk
x24	On your past five commutes to work, how often did you take a motorcycle
x25	On your past five commutes to work, how often did you take a mode other than those listed in variables 18 through 24
x26	On your past five commutes to work, how often have you changed route or departure time
x27	HOV lanes save all commuters time: 0 if strongly disagree, 1 if disagree, 2 if neutral, 3 if agree, 4 if agree strongly
x28	Existing HOV lanes are being adequately used: 0 if strongly disagree, 1 if disagree, 2 if neutral, 3 if agree, 4 if agree strongly
x29	HOV lanes should be open to all traffic: 0 if strongly disagree, 1 if disagree, 2 if neutral, 3 if agree, 4 if agree strongly
x30	Converting some regular lanes to HOV lanes is a good idea: 0 if strongly disagree, 1 if disagree, 2 if neutral, 3 if agree, 4 if agree strongly
x31	Converting some regular lanes to HOV lanes is a good idea only if it is done before traffic congestion becomes serious: 0 if strongly disagree, 1 if disagree, 2 if neutral, 3 if agree, 4 if agree strongly
x32	Gender: 1 if male, 0 if female
x33	Age in years: 0 if under 21, 1 if 22 to 30, 2 if 31 to 40, 3 if 41 to 50, 4 if 51 to 64, 5 if 65 or greater

x34	Annual household income (US dollars per year): 0 if no income, 1 if 1 to 9,999, 2 if 10,000 to 19,999, 3 if 20,000 to 29,999, 4 if 30,000 to 39,999, 5 if 40,000 to 49,999, 6 if 50,000 to 74,999, 7 if 75,000 to 100,000, 8 if over 100,000
x35	Highest level of education: 0 if did not finish high school, 1 if high school, 2 if community college or trade school, 3 if college/university, 4 if post college graduate degree
x36	Number of household members
x37	Number of adults in household (aged 16 or more)
x38	Number of household members working outside the home
x39	Number of licensed motor vehicles in the household
x40	Postal zip code of work place
x41	Postal zip code of home
x42	Type of survey comment left by respondent regarding opinions on HOV lanes: 0 if no comment on HOV lanes, 1 if comment not in favor of HOV lanes, 2 comment positive toward HOV lanes but critical of HOV lane policies, 3 comment positive toward HOV lanes, 4 neutral HOV lane comment

```
read;nvar=42;nobs=250;file=D:\old_drive_d\new_laptop\CE697N-disk\SURVEYS-L-BP.csv$
create;if(x1=0)dalone=1$
create;if(x33>3&x32=1)oldmen=1$
create;if(x35>2)college=1$

RECODE;x27;0,1=1;2,3,4=0$
RECODE;x28;0,1=1;2,3,4=0$
RECODE;x29;0,1=1;2,3,4=0$
RECODE;x30;0,1=1;2,3,4=0$
RECODE;x31;0,1=1;2,3,4=0$

--> mprobit;lhs=x27,x28,x29,x30,x31
    ;eq1=one,dalone,oldmen
    ;eq2=one,dalone,oldmen
    ;eq3=one,dalone,x8,oldmen
    ;eq4=one,dalone,x37
    ;eq5=one,oldmen,college
    ;marginal effects$
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Multivariate Probit Model:  5 equations.    |
| Maximum Likelihood Estimates                |
| Model estimated: Feb 18, 2015 at 10:51:20AM.|
| Dependent variable             MVProbit     |
| Weighting variable                 None     |
| Number of observations              250     |
| Iterations completed                 35     |
```

Log likelihood function	-688.7882
Number of parameters	26
Info. Criterion: AIC =	5.71831
Finite Sample: AIC =	5.74349
Info. Criterion: BIC =	6.08454
Info. Criterion:HQIC =	5.86570
Replications for simulated probs. =	100

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Index function for X27				
Constant	-.32157777	.19026314	-1.690	.0910	
DALONE	.47332787	.21173995	2.235	.0254	.77200000
OLDMEN	-.15607381	.24811487	-.629	.5293	.13600000
	Index function for X28				
Constant	-.01099928	.18750087	-.059	.9532	
DALONE	.63190760	.21443448	2.947	.0032	.77200000
OLDMEN	.42116517	.26681884	1.578	.1145	.13600000
	Index function for X29				
Constant	.89729463	.22564543	3.977	.0001	
DALONE	-.93431181	.24224050	-3.857	.0001	.77200000
X8	-.00037643	.00062377	-.603	.5462	-11.5120000
OLDMEN	-.35167770	.24867900	-1.414	.1573	.13600000
	Index function for X30				
Constant	-.34260970	.28486650	-1.203	.2291	
DALONE	.66066409	.22129438	2.985	.0028	.77200000
X37	-.12221300	.09606297	-1.272	.2033	2.16000000
	Index function for X31				
Constant	-.08184549	.16687673	-.490	.6238	
OLDMEN	.33819408	.23984110	1.410	.1585	.13600000
COLLEGE	-.28439965	.18246065	-1.559	.1191	.78400000
	Correlation coefficients				
R(01,02)	.65146405	.08010881	8.132	.0000	
R(01,03)	-.68827485	.07439357	-9.252	.0000	
R(02,03)	-.68014504	.08069851	-8.428	.0000	
R(01,04)	.49820795	.09007825	5.531	.0000	
R(02,04)	.48659862	.09765957	4.983	.0000	
R(03,04)	-.51128771	.09012354	-5.673	.0000	
R(01,05)	.45454172	.09626375	4.722	.0000	
R(02,05)	.33275375	.11143455	2.986	.0028	
R(03,05)	-.27543541	.10974492	-2.510	.0121	
R(04,05)	.63744041	.07616090	8.370	.0000	

Partials of E[y1|other vars=1,X] wrt X
Computed at the means of all RHS vars.
Conditional mean is Prob[X27 =1] given
X28 through X31 all equal 1.000.
Estimate of conditional mean = .54350

Variable	Mean of Variable	Coefficient in Equation X27	X28	X29	X30	X31	Marginal Effect
ONE	1.00000	-.32158	-.01100	.89729	-.34261	-.08185	.00000
DALONE	.77200	.47333	.63191	-.93431	.66066	.00000	-.03988
OLDMEN	.13600	-.15607	.42117	-.35168	.00000	.33819	-.20510
X8	-11.5120	.00000	.00000	-.00038	.00000	.00000	-.00009
X37	2.16000	.00000	.00000	.00000	-.12221	.00000	.00385
COLLEGE	.78400	.00000	.00000	.00000	.00000	-.28440	.01394