Queensland University of Technology
Transport Data Analysis and Modeling Methodologies

Lab Session #9
(Ordered Probit with Random Effects)

A survey of 56 subjects was conducted on freeways in the Seattle area (see text pages 352-357). Each subject drove a vehicle over 40 freeway segments (thus each subject can generate as many as 40 observations if there is no missing data). As they drove over the test segments, they were asked: "How would you rank the roughness of the road on a scale from one to five – with one being the smoothest (or the best) and five being the roughest (or the worst)?" Data were collected on the type of vehicle being used (minivan, pickup, etc.), in-vehicle-cabin noise (dBA), vehicle speed (km/h), socioeconomic information, IRI measurement, age of the roadway surface, information on patching, and the Pavement Structural Condition (PSC). This last term is calculated separately for flexible and rigid pavements based on the amount and severity of various distresses and its values range from 100 (excellent pavement condition) to zero (completely deteriorated pavement).

Your task is to estimate a model of the ordered response of roughness perception while accounting for repeat observations from individual subjects:

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. (e.g. 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 pavement-pds.txt):

Variable Number	Explanation
1	Individual number
2	Roadway segment number
3	pds for Limdep random effects
4	Ruffness ranking: 1 = very smooth; 5 = very rough
5	Sedan: 1 if yes, 0 if no
6	Sport utility vehicle: 1 if yes, 0 if no
7	Pickup: 1 if yes, 0 if no
8	Mininvan: 1 if yes, 0 if no
9	Nosie dBA reading
10	Speed in miles per hour
11	Level of service: A=1, B=2, C=3,D=4,E=5,F=6
12	User regularly uses I-5: 1 if yes, 0 if no
13	User regularly uses I-90: 1 if yes, 0 if no
14	User regularly uses I-405: 1 if yes, 0 if no
15	User regularly uses SR-520: 1 if yes, 0 if no
16	Female: 1 if yes, 0 if no
17	Married: 1 if yes, 0 if no
18	Age: 0 = Less than 21; 1 = 21 - 25; 2 = 26-30; 3=31-35; 4 = 36 - 40; 5 = 41 - 45; 6 = 46 - 50; 7 = 51 - 55; 8 = 56 - 60; 9 = 61 - 65; 10 = 66 - 70; 11 = Over 70
19	Income: 0 = no income; 1 = under $15,000; 2 = $15,000 - $24,999; 3 = $25,000 - $34,999; 4 = $35,000 - $44,999; 5 = $45,000 - $54,999; 6 = $55,000 - $64,999; 7 = $65,000 - $74,999; 8 = $75,000 - $84,999; 9 = $85,000 - $99,999; 10 = $100,000 - $150,000; 11 = over $150,000

20	Education: 1 = some high school; 2 = high school diploma; 3 = technical college degree (AA); 4 = college degree (BS or BA) 5 = post-graduate degree
21	Vehicle type normally driven: (miscoded, do not use)
22	Number of household vehciles
23	Household size
24	Number of household infants
25	Number of household children
26	Number of workers
27	International roughness index (IRI) in m/km
28	Roadway surface age
29	Visible wear: 1 if yes, 0 if no
30	Visible joints: 1 if yes, 0 if no
31	Visible patching: 1 if yes, 0 if no
32	Bridge in section: 1 if yes, 0 if no
33	Surface type: 1 if concrete, 0 if asphalt
34	Rut depth in mm
35	Pavement structural condition index (PSC)
36	Section lenght in miles
37	Number of lanes
38	Cracking present: 1 if yes, 0 if no
39	Scaling present: 1 if yes, 0 if no
40	Faulting present: 1 if yes, 0 if no
41	Spalling present: 1 if yes, 0 if no
42	IRI change from last section (m/km)
43	Nosie change from last section (dBA)

```
--> RESET
--> read;nvar=43;nobs=2240;file=D:\old_drive_d\new_laptop\CE697M\pavement-pds.txt
--> histogram;rhs=x4$
--> create;if(x4=1)xx4=0$
--> create;if(x4=2)xx4=1$
--> create;if(x4=3)xx4=2$
--> create;if(x4=4)xx4=3$
--> create;if(x4=5)xx4=4$
--> reject;x4=-999$
--> ordered;lhs=xx4;rhs=one,x9,x16,x27,x28
    ;pds=x3;margin$
```

```
+---------------------------------------------------------------------+
| Dependent variable is binary, y=0 or y not equal 0                  |
| Ordinary    least squares regression    Weighting variable = none   |
| Dep. var. = Y=0/Not0 Mean=   .8421294172    , S.D.=   .3647033082   |
| Model size: Observations =    2179, Parameters =   5, Deg.Fr.=   2174 |
| Residuals:  Sum of squares= .3251072901D+04, Std.Dev.=       1.22288 |
| Fit:        R-squared=*********, Adjusted R-squared =      -10.24314 |
| Diagnostic: Log-L =  -3527.7968, Restricted(b=0) Log-L =    -893.4726 |
|             LogAmemiyaPrCrt.=     .405, Akaike Info. Crt.=      3.243 |
+---------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant -.5531567899E-01  .79501077E-01    -.696   .4866
 X9        .4853153867E-03  .50353896E-03     .964   .3351  85.693896
 X16       .1085572574      .53419718E-01    2.032   .0421  .40339605
 X27       .5682371126      .46826433E-01   12.135   .0000  1.9368380
 X28      -.1590470840E-01  .27855013E-02   -5.710   .0000  18.140431

Normal exit from iterations. Exit status=0.
```

```
+-------------------------------------------+
| Ordered Probit Model                      |
| Maximum Likelihood Estimates              |
| Dependent variable                   XX4  |
| Weighting variable                   ONE  |
| Number of observations              2179  |
| Iterations completed                  14  |
| Log likelihood function       -2645.567   |
| Restricted log likelihood     -3187.274   |
| Chi-squared                    1083.414   |
| Degrees of freedom                    4   |
| Significance level             .0000000   |
|    Cell frequencies for outcomes          |
|  Y Count Freq  Y Count Freq  Y Count Freq |
|  0   344 .157  1   769 .352  2   601 .275 |
|  3   351 .161  4   114 .052               |
+-------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Index function for probability				
Constant	-.4087697743	.83223827E-01	-4.912	.0000	
X9	.1187111410E-02	.60834679E-03	1.951	.0510	85.693896
X16	-.1662863090	.46908654E-01	-3.545	.0004	.40339605
X27	.7277308874	.44193328E-01	16.467	.0000	1.9368380
X28	.1642796558E-01	.23191594E-02	7.084	.0000	18.140431
	Threshold parameters for index				
Mu(1)	1.310808503	.40469055E-01	32.390	.0000	
Mu(2)	2.411947498	.53429323E-01	45.143	.0000	
Mu(3)	3.548485694	.70158413E-01	50.578	.0000	

Normal exit from iterations. Exit status=0.

RANDOM EFFECTS Ordered Probit Model	
Maximum Likelihood Estimates	
Dependent variable	XX4
Weighting variable	ONE
Number of observations	2179
Iterations completed	26
Log likelihood function	-2389.480
Restricted log likelihood	-2645.567
Chi-squared	512.1733
Degrees of freedom	1
Significance level	.0000000
Unbalanced panel has	56 individuals.

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Index function for probability				
Constant	-.2730483606	.18302227	-1.492	.1357	
X9	.5625783440E-03	.17690875E-02	.318	.7505	85.693896
X16	-.5020315727	.84405249E-01	-5.948	.0000	.40339605
X27	.8743517479	.50596167E-01	17.281	.0000	1.9368380
X28	.1932085053E-01	.31342059E-02	6.165	.0000	18.140431
	Threshold parameters for index model				
Mu(01)	1.561271285	.39183897E-01	39.845	.0000	
Mu(02)	2.861242547	.36358293E-01	78.696	.0000	
Mu(03)	4.221001531	.43737096E-01	96.508	.0000	
	Std. Deviation of random effect				
Sigma	.7054694476	.42093354E-01	16.760	.0000	

Marginal Effects for OrdProbt				
Variable	XX4=0	XX4=1	XX4=2	XX4=3
ONE	.0295	.0793	-.0586	-.0465
X9	-.0001	-.0002	.0001	.0001
X16	.0542	.1458	-.1077	-.0856
X27	-.0944	-.2539	.1875	.1490
X28	-.0021	-.0056	.0041	.0033