PURDUE
UNIVERSITY

School of Civil Engineering

CE697M – Statistical and Econometric Methods II

Assignment #7
(Mixed Logit Analysis)

You are given data accident, evirnomental, traffic, and roadway geometric data from 275 segments of highway in Washington State. The data compile accidents per year for 5 years from 1990 to 1994. Thus there are a total of 1375 observations (275 x 5). The injury data consist of three possible outcomes: no injury, possible injury, injury. Your task is to estimate a mixed logit model of these three possible discrete outcomes.

The mixed logit model allows for parameter variations across roadway segments (i.e. variations in $\boldsymbol{\beta}$), a mixing distribution is introduced giving injury-severity proportions (see Train 2003),

$$P_{in} = \int \frac{EXP\left[\boldsymbol{\beta}_i X_{in}\right]}{\sum_I EXP\left[\boldsymbol{\beta}_i X_{In}\right]} f\left(\boldsymbol{\beta} \mid \varphi\right) d\boldsymbol{\beta}$$

where $f(\boldsymbol{\beta} \mid \boldsymbol{\varphi})$ is the density function of $\boldsymbol{\beta}$ with $\boldsymbol{\varphi}$ referring to a vector of parameters of the density function (mean and variance), and all other terms are as previously defined. Equation 3 is the formulation for the mixed logit model. For model estimation, $\boldsymbol{\beta}$ can now account for segment-specific variations of the effect of $\boldsymbol{X}$ on injury-severity proportions, with the density function $f(\boldsymbol{\beta} \mid \boldsymbol{\varphi})$ used to determine $\boldsymbol{\beta}$. Mixed logit proportions are then a weighted average for different values of $\boldsymbol{\beta}$ across roadway segments where some elements of the vector $\boldsymbol{\beta}$ may be fixed and some may randomly distributed. If the parameters are random, the mixed logit weights are determined by the density function $f(\boldsymbol{\beta} \mid \boldsymbol{\varphi})$. Most studies have used a continuous form of this density function in model estimation (such as a normal distribution) and this is what you are to use.

In your specification, consider random variable possibilities including constant or fixed (C), normally distributed (N) and log-normally distributed (L).

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 (the theory behind the inclusion of your selected variables). Include *t*-statistics and justify the signs of your variables.

Variables available for your specification are (in file Before-Data-nlogit-r.csv):

Variable Number	Explanation
ID	Segment ID number
ROUTE	Route Number
LENGTH	Segment length in miles
INCLANES	Number of lanes in increasing milepost direction
DECLANES	Number of lanes in decreasing milepost direction
WIDTH	Total combined width of all lanes
MIMEDSH	Minimum median shoulder in feet
MXMEDSH	Maximum median shoulder in feet
SPEED	Speed limit
URB	Indicates urban area (1=yes, 0=no)
FC	Functional class (1=local, 2=collector, 3=arterial, 4=principal arterial, 5=interstate)
AADT	Average Annual Daily Traffic
SINGLE	Daily percentage of single unit trucks
DOUBLE	Daily percentage of tractor and trailer trucks
TRAIN	Daily percentage of tractor and two-trailer trucks
PEAKHR	Percent of daily traffic in the peak hour
GRADEBR	Number of grade breaks in the segment
MIGRADE	Minimum grade in the segment
MXGRADE	Maximum grade in the segment
MXGRDIFF	Maximum grade difference in the segment
TANGENT	Tangent length in the segment

CUMTAN	Cumulative tangent length in the segment
CURVES	Number of cureves in the segment
MINRAD	Minimum radius in feet
ACCESS	Segment access control (0=none, 1=partial, 3=full)
ACCYR	Accident year (0=1990, 1=1991, 2=1992, 3=1993, 4=1994)
MEDWIDTH	Median width (1=less than 30ft; 2=30 to 40ft; 3=40 to 50ft; 4=50 to 60ft to 5=high)
PRECIP	Indicates presence of a precipitation (1=yes; 0=no)
FRICTION	Friction value (0 to 100 with 100 being high)
ADTLANE	Average daily travel per lane
SLOPE	Segment slope (0=flat, 1=slight, 2=medium, 3=high)
ELEVDIFF	Elevation difference in the segment in feet
VEGET	Indicates presence of vegitation in the clear zone (1=yes; 0=no)
OTHEROBS	Indicates presence of other obstructions in the clear zone (1=yes; 0=no)
DITCH	Indicates presence of a ditch (1=yes; 0=no)
PILLAR	Indicates presence of a pillars in clear zone (1=yes; 0=no)
GRAVSHD	Indicates gravel shoulder (1=yes; 0=no)
CROSOVER	Indicates presence of a median crossover (1=yes; 0=no)
INTECHAG	Indicates number of interganges in the segment
EXITENTR	Indicates presence of exit/entrance ramps (1=yes; 0=no)
AVEPRE	Average precipitation per month in inches
AVESNOW	Average snowfall per month in inches
INJFREQ	Injury frequency (dependent variable – property damage only, possible injury, injury)

```
read;nvar=43;nobs=4125;names=
ID,ROUTE,LENGTH,INCLANES,DECLANES,WIDTH,MIMEDSH,MXMEDSH,SPEED,URB,FC,AADT,
SINGLE,DOUBLE,TRAIN,PEAKHR,GRADEBR,MIGRADE,MXGRADE,MXGRDIFF,
TANGENT,CUMTAN,CURVES,MINRAD,ACCESS,ACCYR,MEDWIDTH,
PRECIP,FRICTION,ADTLANE,SLOPE,ELEVDIFF,VEGET,OTHEROBS,DITCH,
PILLAR,GRAVSHD,CROSOVER,INTECHAG,EXITENTR,AVEPRE,AVESNOW,INJFREQ;
FILE="D:\OLD_DRIVE_D\NEW_LAPTOP\CE697M\MIXED-LOGIT-R.CSV"$

dstat;rhs=injfreq$

/*ADT VARIABLE*/
create;laneadt=aadt/(inclanes+declanes)$
create;lnlanadt=log(laneadt)$
create;lnaadt=log(aadt)$
create;density=laneadt/length$

/*FRICTION VARIABLE*/
create;if(friction<=30)             lowfri=1$
create;if(friction>30&friction<50)  medfri=1$
create;if(friction>=50)              hifri=1$

/*CURVE/LENGTH VARIABLE*/
/*dstat;rhs=curvmile$*/
create;curvmile=curves/length$
create;if(curvmile<=0.5)               lowcvmil=1;(else)lowcvmil=0$
create;if(curvmile>0.5&curvmile<=2.5) medcvmil=1;(else)medcvmil=0$
create;if(curvmile>2.5)                hicvmil =1;(else)hicvmil=0$
/*%TRUCK VARIABLE*/
create;truck=single+double+train$
create;pcttruck=truck/aadt$

/*MEDWIDTH VARIABLE*/
/*dstat;rhs=medwidth$
histogram;rhs=medwidth$*/
create;if(medwidth=1)med030=1$
create;if(medwidth=2)med3040=1$
create;if(medwidth=3)med4050=1$
create;if(medwidth=4)med5060=1$
create;if(medwidth=5)med60=1$

/*SPEED VARIABLE*/
/*dstat;rhs=speed$
histogram;rhs=speed$*/
create;if(speed<=50)speed1=1$
create;if(speed<=55)speed2=1$
create;if(speed>55)speed3=1$
create;if(speed>=55)speed4=1$

/*FUNCTIONAL CLASS VARIABLE*/
/*dstat;rhs=fc$
histogram;rhs=fc$*/
create;if(fc=1)local=1$
create;if(fc=5)intstate=1$

/*ACESS VARIABLE*/
create;if(access=0)none   =1$
create;if(access=1)partial=1$
create;if(access=2)full   =1$

/*SLOPE VARIABLE*/
create;if(slope=0)flat=1$
create;if(slope=1)slight=1$
create;if(slope=2)medium=1$
create;if(slope=0 |slope= 1)slpflat=1;(else)slpflat=0$
create;if(slope=2)slpmed=1;(else)slpmed=0$
```

```
/***WEATHER VARIABLE***/

/*histogram;rhs=avepre,avesnow$*/

create;if(avepre<=1.5)lowpre=1;(else)lowpre=0$
create;if(avepre>1.5&avepre<=2.5)medpre=1;(else)medpre=0$
create;if(avepre>2.5)hipre=1;(else)hipre=0$
*dstat;rhs=avepre$
histogram;rhs=avepre$*/
create;if(avesnow<=1)norsnow=1$
create;if(avesnow>1)hisnow=1$

create;lanewid=(inclanes+declanes)/width$
dstat;rhs=lanewid$
create;if(lanewid<12)nlanwid=1;(else)nlanwid=0$
create;if(lanewid>12)wlanwid=1;(else)wlanwid=0$



--> nlogit;lhs=injfreq;
    choices=pdo,pinj,inj;
    model:
    U(pdo)=a0+a1*laneadt+a2*avesnow/
    U(pinj)=b0+b2*pcttruck/
    U(inj)=c1*curvmile+c2*truck+c3*gradebr
    ;fcn=a0(N),a1(N),a2(N),
    b0(N),b2(N),c1(C),c2(N),c3(C);rpl;frequencies;parameter;pts=100,R;
    effects:laneadt(*)/avesnow(*)/truck(*)/curvmile(*)/truck(*)/gradebr(*)$
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Start values obtained using nonnested model |
| Maximum Likelihood Estimates                |
| Model estimated: Apr 12, 2007 at 08:58:28AM.|
| Dependent variable               Choice     |
| Weighting variable                 None     |
| Number of observations             1280     |
| Iterations completed                  5     |
| Log likelihood function       -22124.43     |
| R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj |
| No coefficients -24849.5114  .10966  .10477 |
| Constants only.  Must be computed directly. |
|                  Use NLOGIT ;...; RHS=ONE $ |
| Chi-squared[ 6]           =    583.61703    |
| Prob [ chi squared > value ] =   .00000     |
| Response data are given as frequencies.     |
| Number of obs.=  1375, skipped  95 bad obs. |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
 A0            1.07678517      .07142386    15.076   .0000
 A1          -.799452D-06    .246490D-05     -.324   .7457
 A2             .02534778      .00807067     3.141   .0017
 B0             .36442718      .05256625     6.933   .0000
 B2          -455.074899     33.3621568   -13.640   .0000
 C1            -.10697396      .02322244    -4.606   .0000
 C2             .01971519      .00298645     6.602   .0000
 C3             .00906180      .00289508     3.130   .0017
```

```
Normal exit from iterations. Exit status=0.
+---------------------------------------------+
| Random Parameters Logit Model               |
| Maximum Likelihood Estimates                |
| Model estimated: Apr 12, 2007 at 09:00:32AM.|
| Dependent variable              INJFREQ     |
| Weighting variable                 None     |
| Number of observations             4125     |
| Iterations completed                 46     |
| Log likelihood function       -22044.71     |
| Restricted log likelihood     -24849.51     |
| Chi squared                    5609.609     |
| Degrees of freedom                   14     |
| Prob[ChiSqd > value] =          .0000000    |
| R2=1-LogL/LogL*  Log-L fncn  R-sqrd  RsqAdj |
| No coefficients -24849.5114  .11287  .10799 |
| Constants only.  Must be computed directly. |
|                  Use NLOGIT ;...; RHS=ONE $ |
| At start values -22124.4350  .00360 -.00188 |
| Response data are given as frequencies.     |
+---------------------------------------------+

+---------------------------------------------+
| Random Parameters Logit Model               |
| Replications for simulated probs. = 100     |
| Number of obs.=  1375, skipped  95 bad obs. |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] |
+---------+--------------+----------------+--------+---------+
         Random parameters in utility functions
 A0            .58777906      .08333468     7.053   .0000
 A1          .201631D-04    .317441D-05     6.352   .0000
 A2            .03879625      .01242345     3.123   .0018
 B0            .54142327      .06656279     8.134   .0000
 B2          -3064.56522     400.578896    -7.650   .0000
 C1           -.14567488      .02398017    -6.075   .0000
 C2            .00228821      .00368193      .621   .5343
 C3            .00797700      .00309501     2.577   .0100
         Derived standard deviations of parameter distributions
 NsA0          .04533835      .16083801      .282   .7780
 NsA1        .514171D-05    .156966D-04      .328   .7432
 NsA2          .10001307      .03846814     2.600   .0093
 NsB0          .27696429      .22180600     1.249   .2118
 NsB2         2764.01180     368.497427     7.501   .0000
 CsC1             .000000   .....(Fixed Parameter).......
 NsC2          .01441151      .00945180     1.525   .1273
 CsC3            .000000   .....(Fixed Parameter).......
```