Transport Data Analysis and Modeling Methodologies

Lab Session #3
(Count Data – Negative Binomial Regression)

You are given accident, evirnomental, traffic, and roadway geometric data from 275 segments of highway in Washington State. The data are from 1990. Your task is to estimate a count-data model for the total number of accidents on these segments.

The data are going to be overdispersed, so a negative binomial regression will likely be statistically superior to a Poisson regression. As with the Poisson regression, you will be estimating a parameter vector β such that:

$$\lambda = EXP(\beta X)$$

where λ is the Poisson parameter that in this case is the expected number of departure changes per week.

In your analysis please provide the following:

1. Provide the results of your best model specification and compute ρ^2 (see equation 11.12 in the text).

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 Ex16-2.txt):

Variable Number	**Explanation**
ID	Segment ID number
FREQ	Number of accidents
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 (mi/h)
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
CURVES	Number of cureves in the segment

MINRAD	Minimum radius in feet
ACCESS	Segment access control (0=none, 1=partial, 3=full)
MEDWIDTH	Median width (1=less than 30ft; 2=30 to 40ft; 3=40 to 50ft; 4=50 to 60ft to 5=high)
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)
INTECHAG	Indicates number of interganges in the segment
AVEPRE	Average precipitation per month in inches
AVESNOW	Average snowfall per month in inches

```
--> read;nvar=32;nobs=275;names=ID,FREQ,ROUTE,LENGTH,INCLANES,DECLANES,WIDTH,
    MIMEDSH,MXMEDSH,SPEED,URB,FC,AADT,SINGLE,DOUBLE,TRAIN,PEAKHR,GRADEBR,
    MIGRADE,MXGRADE,MXGRDIFF,TANGENT,CURVES,MINRAD,ACCESS,MEDWIDTH,
    FRICTION,ADTLANE,SLOPE,INTECHAG,AVEPRE,AVESNOW
    ;FILE=D:/old_drive_d/book/book2e-Data/Ex16-2.txt$
--> create;expose=aadt*length*365/100000000$
--> negbin;lhs=freq;rhs=one,expose
    ;marginal effects$
+--------------------------------------------+
| Poisson Regression                         |
| Maximum Likelihood Estimates               |
| Model estimated: May 18, 2015 at 00:47:07PM.|
| Dependent variable                FREQ     |
| Weighting variable                None     |
| Number of observations             275     |
| Iterations completed                 6     |
| Log likelihood function       -2150.001     |
| Number of parameters                 2     |
| Info. Criterion: AIC =         15.65091     |
|   Finite Sample: AIC =         15.65108     |
| Info. Criterion: BIC =         15.67722     |
| Info. Criterion:HQIC =         15.66147     |
| Restricted log likelihood     -3178.572     |
| McFadden Pseudo R-squared      .3235954     |
| Chi squared                   2057.143     |
| Degrees of freedom                   1     |
| Prob[ChiSqd > value] =         .0000000     |
+--------------------------------------------+

+--------------------------------------------+
| Poisson Regression                         |
| Chi- squared =  3677.55492  RsqP=   .5055  |
| G  - squared =  3217.09201  RsqD=   .3900  |
| Overdispersion tests: g=mu(i)  :  4.205    |
| Overdispersion tests: g=mu(i)^2:  4.699    |
+--------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    2.41799169       .01804768   133.978   .0000
 EXPOSE  |    1.11300618       .01856630    59.948   .0000    .25784008

Normal exit from iterations. Exit status=0.

+--------------------------------------------+
| Negative Binomial Regression               |
| Maximum Likelihood Estimates               |
| Model estimated: May 18, 2015 at 00:47:07PM.|
| Dependent variable                FREQ     |
| Weighting variable                None     |
| Number of observations             275     |
| Iterations completed                 8     |
| Log likelihood function       -988.1926     |
| Number of parameters                 3     |
| Info. Criterion: AIC =          7.20867     |
|   Finite Sample: AIC =          7.20900     |
| Info. Criterion: BIC =          7.24813     |
| Info. Criterion:HQIC =          7.22451     |
| Restricted log likelihood     -2150.001     |
| McFadden Pseudo R-squared      .5403757     |
| Chi squared                   2323.616     |
| Degrees of freedom                   1     |
| Prob[ChiSqd > value] =         .0000000     |
| NegBin form 2; Psi(i) = theta              |
+--------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	1.97160174	.06760221	29.165	.0000	
EXPOSE	2.36683951	.18508989	12.788	.0000	.25784008
Dispersion parameter for count data model					
Alpha	.68441177	.05017018	13.642	.0000	

Partial derivatives of expected val. with respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point 53.3849
Scale Factor for Marginal Effects 53.3849

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	105.253692	541.739365	.194	.8460	
EXPOSE	126.353407	778.991947	.162	.8711	.25784008