Queensland University of Technology

Transport Data Analysis and Modeling Methodologies

Lab Session #2
(Count Data – Poisson Regression)

You are given 204 observations from a travel survey conducted in the Seattle metropolitan area. The purpose of the survey was to study the number of times (per week) commuters' changed their departure time on their work-to-home trip to avoid traffic congestion. The data are non-negative integers with the mean approximately equal to the variance thus the data are well suited to the Poisson regression approach. Remember in a Poisson regression, you are 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. Manually compute the initial log-likelihood (see equation 11.4 in the text)

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

3. 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.

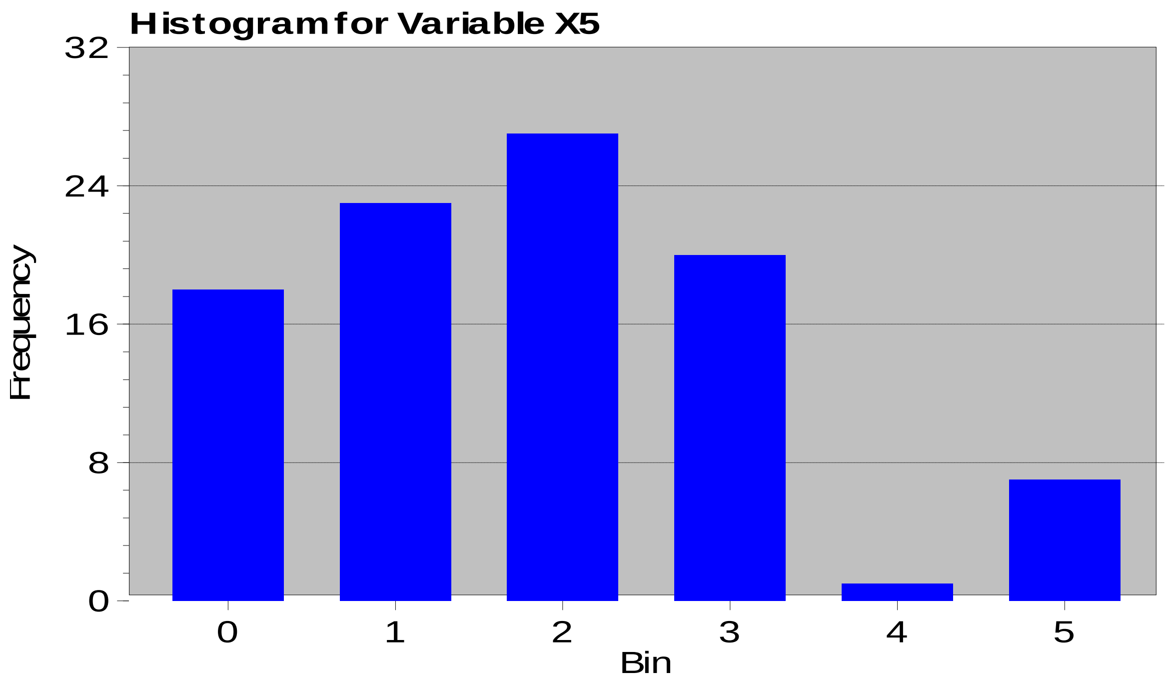
4. After you have your best model, run a negative binomial model (using the "negbin" command instead of the "poisson" command) to ensure that the negative binomial overdispersion parameter is not significantly different from zero.

Variables available for your specification are (file *tobit.dat*)

Variable Number	Explanation
x1	Household number
x2	Do you ever delay work-to-home departure to avoid traffic congestion? 1-yes, 0-no
x3	If sometimes delay, on average how many minutes do you delay?
x4	If sometimes delay, do you 1-perform additional work, 2-engage in non-work activities, or 3-do both?
x5	If sometimes delay, how many times have you delayed in the past week?
x6	Mode of transportation used work-to-home: 1-car SOV, 2-carpool, 3-vanpool, 4-bus, 5 other.
x7	Primary route (work-to-home): 1-I90, 2-I5, 3-SR520, 4-I405, 5-other
x8	Do you generally encounter traffic congestion on you work-to-home trip? 1-yes, 0-no
x9	Age: 1-(<25), 2-(26-30), 3-(31-35), 4-(36-40), 5-(41-45), 6-(46-50), 7-(>50)
x10	Gender: 1-male, 0-female
x11	Number of cars in household
x12	Number of children in household
x13	Income: 1 - less than 20000, 2 - 20000 to 29999, 3 - 30000 to 39999, 4 - 40000 to 49999, 5 - 50000 to 59999, 6 - >60000
x14	Do you have flexible work hours? 1-yes, 0-no
x15	Distance from work to home (in miles)
x16	Face LOS D or worse? 1-yes, 0-no
x17	Ratio of actual travel time to free-flow travel time
x18	Population of work zone
x19	Retail employment in work zone
x20	Service employment in work zone
x21	Size of work zone (in acres)

```
--> RESET
--> read;nvar=21;nobs=204;file=D:\old_drive_d\new_laptop\CE697N-disk\TOBIT.DAT$
--> reject;x2=0$
--> create;if(x7=3)sr520=1$
--> create;if(x7=2)I5=1$
--> dstat;rhs=x5$
                              Descriptive Statistics
               All results based on nonmissing observations.
Variable          Mean           Std.Dev.         Minimum          Maximum      Cases
-------------------------------------------------------------------------------------
X5         .183333333D+01  .137394297D+01  .000000000D+00  .500000000D+01      96

--> histogram;rhs=x5$
Histogram for X5            NOBS=     96, Too low:     0, Too high:     0
Bin  Lower limit    Upper limit       Frequency     Cumulative Frequency
========================================================================
  0         .000          1.000         18  ( .1875)          18( .1875)
  1        1.000          2.000         23  ( .2396)          41( .4271)
  2        2.000          3.000         27  ( .2813)          68( .7083)
  3        3.000          4.000         20  ( .2083)          88( .9167)
  4        4.000          5.000          1  ( .0104)          89( .9271)
  5        5.000          6.000          7  ( .0729)          96(1.0000)
```




```
--> poisson;lhs=x5;rhs=one,sr520,i5,x10,x11,x14,x15,x17;
    limit=6;truncation;upper;marginal effects$

              +---------------------------------------------+
              | Poisson Regression                          |
              | Maximum Likelihood Estimates                |
              | Dependent variable                     X5   |
              | Weighting variable                    ONE   |
              | Number of observations                 96   |
              | Iterations completed                    6   |
              | Log likelihood function       -151.3086     |
              | Restricted log likelihood     -160.5608     |
              | Chi-squared                    18.50428     |
              | Degrees of freedom                      7   |
              | Significance level             .9890572E-02 |
              | RIGHT Truncated data, at Y =  5.            |
              | Chi- squared =    90.40988  RsqP=   .0757   |
              | G  - squared =   103.41426  RsqD=   .1168   |
              +---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	1.821166565	.75186875	2.422	.0154	
SR520	-.5376391967	.25629632	-2.098	.0359	.16666667
I5	-.3251529733	.18668242	-1.742	.0816	.34375000
X10	-.4854679467E-01	.17966580	-.270	.7870	.69791667
X11	-.1210107221	.87127154E-01	-1.389	.1649	1.8854167
X14	-.3929141620	.17063688	-2.303	.0213	.63541667
X15	-.2822605374E-01	.20647975E-01	-1.367	.1716	7.7083333
X17	-.1453901223	.30066940	-.484	.6287	1.9593750

```
+-------------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
| Conditional Mean at Sample Point   1.8925 |
| Scale Factor for Marginal Effects  1.7468 |
+-------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	3.181282683	1.3838428	2.299	.0215	
SR520	-.9391684974	.47673703	-1.970	.0488	.16666667
I5	-.5679895201	.34069727	-1.667	.0955	.34375000
X10	-.8480337833E-01	.32250420	-.263	.7926	.69791667
X11	-.2113861094	.15790943	-1.339	.1807	1.8854167
X14	-.6863573292	.31341031	-2.190	.0285	.63541667
X15	-.4930633896E-01	.37373581E-01	-1.319	.1871	7.7083333
X17	-.2539729684	.53870794	-.471	.6373	1.9593750

--> negbin;lhs=x5;rhs=one,sr520,i5,x10,x11,x14,x15,x17;
limit=6;truncation;upper;marginal effects$

```
+---------------------------------------------+
| Negative Binomial Regression                |
| Maximum Likelihood Estimates                |
| Dependent variable                    X5    |
| Weighting variable                   ONE    |
| Number of observations                96    |
| Iterations completed                  39    |
| Log likelihood function        -150.4476    |
| Restricted log likelihood      -151.3086    |
| Chi-squared                     1.722039    |
| Degrees of freedom                     1    |
| Significance level              .1894308    |
| RIGHT Truncated data, at Y =  5.            |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	1.833827968	.78699537	2.330	.0198	
SR520	-.5426164710	.27041492	-2.007	.0448	.16666667
I5	-.3266237780	.20548460	-1.590	.1119	.34375000
X10	-.4718568664E-01	.18928631	-.249	.8031	.69791667
X11	-.1224310050	.90114353E-01	-1.359	.1743	1.8854167
X14	-.3948060093	.18398274	-2.146	.0319	.63541667
X15	-.2857723705E-01	.22525965E-01	-1.269	.2046	7.7083333
X17	-.1490800157	.31844293	-.468	.6397	1.9593750
	Overdispersion parameter for negative binomial model				
Alpha	.8980286514E-05	.10931848	.000	.9999	