PURDUE
UNIVERSITY

School of Civil Engineering

CE615 – Statistical and Econometric Methods II

Assignment #5
(Count Data – Zero-Inflated Models)

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' delayed their departure time on their work-to-home trip to avoid traffic congestion. The data are non-negative integers and are thus well suited to the Poisson regression approach. 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.

Recall that in CE697N we used variable X2 (see next page) to eliminate people in the sample that said they never delayed (leaving just 96 observations). In this assignment, use the full 204 observations and explore various zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) possibilities.

Include a discussion of your findings with the various model forms and a select and discuss what you believe to be your best model.

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

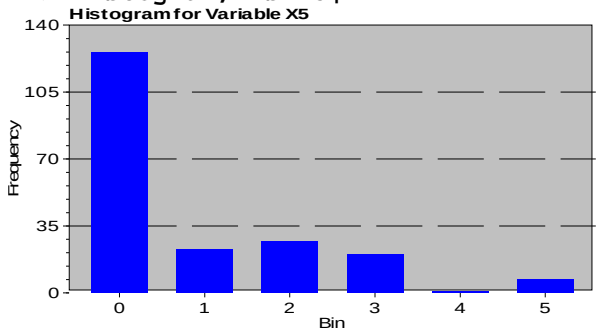
Variable Number	Explanation
x1	Household number
x2	Do you ever delay work-to-home departure to avoid traffic congestion? 1-yes, 0-no
x3	If sometimes delay, on average how many minutes do you delay?
x4	If sometimes delay, do you 1-perform additional work, 2-engage in non-work activities, or 3-do both?
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, 2-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: TOBIT.DAT$
--> create;if(x7=3)sr520=1$
--> create;if(x7=2)I5=1$
--> create;if(x13=6)highinc=1$
--> create;if(x6=1)sov=1$
--> dstat;rhs=x5$
```

Descriptive Statistics
All results based on nonmissing observations.

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases
X5	.862745098D+00	.131335854D+01	.000000000D+00	.500000000D+01	204

```
--> histogram;rhs=x5$
```




```
--> poisson;lhs=x5
    ;rhs=one,sr520,x14,x15,x17
    ;zip
    ;limit=6;truncation;upper$

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

Zero Altered Poisson Regression Model			
Logistic distribution used for splitting model.			
ZAP term in probability is F[tau x ln LAMBDA]			
Comparison of estimated models			
	Pr[0|means]	Number of zeros	Log-likelihood
Poisson	.56606	Act.= 126 Prd.= 115.5	-226.14085
Z.I.Poisson	.46033	Act.= 126 Prd.= 93.9	-199.33319
Note, the ZIP log-likelihood is not directly comparable.			
ZIP model with nonzero Q does not encompass the others.			
Vuong statistic for testing ZIP vs. unaltered model is			8.2473
Distributed as standard normal. A value greater than			
+1.96 favors the zero altered Z.I.Poisson model.			
A value less than -1.96 rejects the ZIP model.			

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Poisson/Negbin regression model				
Constant	-1.209309706	.23735218	-5.095	.0000	
SR520	-.2495955115	.76016669E-01	-3.283	.0010	.13725490
X14	-.5167564012E-01	.55712270E-01	-.928	.3536	.62254902
X15	-.1333964424E-02	.50948915E-02	-.262	.7935	7.1519608
X17	.8695039383	.14837588	5.860	.0000	1.6107843
	Zero inflation model				
Tau	-12.03260123	5.2865738	-2.276	.0228	

```
--> poisson;lhs=x5
    ;rhs=one,sr520,x14,x15,x17
    ;rh2=one,sov
    ;zip
    ;limit=6;truncation;upper$
```

```
 +-----------------------------------------------------------------------+
 | Zero Altered Poisson      Regression Model                            |
 | Logistic distribution used for splitting model.                       |
 | ZAP term in probability is F[tau x Z(i)     ]                         |
 | Comparison of estimated models                                        |
 |             Pr[0|means]       Number of zeros        Log-likelihood   |
 | Poisson          .56606   Act.=   126 Prd.=   115.5      -226.14085   |
 | Z.I.Poisson      .56590   Act.=   126 Prd.=   115.4      -220.81374   |
 | Note, the ZIP log-likelihood is not directly comparable.              |
 | ZIP model with nonzero Q does not encompass the others.               |
 | Vuong statistic for testing ZIP vs. unaltered model is      1.8530    |
 | Distributed as standard normal. A value greater than                  |
 | +1.96 favors the zero altered Z.I.Poisson model.                      |
 | A value less than -1.96 rejects the ZIP model.                        |
 +-----------------------------------------------------------------------+
+---------+-------------+----------------+--------+---------+----------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+----------------+--------+---------+----------+
         Poisson/Negbin regression model
 Constant -3.492857734      .41471885       -8.422   .0000
 SR520    -.3855394991      .22534217       -1.711   .0871  .13725490
 X14      -.2250906578      .14615471       -1.540   .1235  .62254902
 X15       .3696268779E-01  .17421499E-01    2.122   .0339  7.1519608
 X17       1.912795285      .20605998        9.283   .0000  1.6107843
         Zero inflation model
 Constant -.3772053113      .46586747        -.810   .4181
 SOV      -2.324911731      1.4480426       -1.606   .1084  .59313725

--> poisson;lhs=x5
    ;rhs=one,sr520,x14,x15,x17
    ;zip
    ;model=N
    ;limit=6;truncation;upper$

 +-----------------------------------------------------------------------+
 | Zero Altered Neg.Binomial Regression Model                            |
 | Logistic distribution used for splitting model.                       |
 | ZAP term in probability is F[tau x ln LAMBDA]                         |
 | Comparison of estimated models                                        |
 |             Pr[0|means]       Number of zeros        Log-likelihood   |
 | Poisson          .56606   Act.=   126 Prd.=   115.5      -226.14085   |
 | Neg. Bin.        .61350   Act.=   126 Prd.=   125.2      -220.68452   |
 | Z.I.Neg_Bin      .42579   Act.=   126 Prd.=    86.9      -198.31463   |
 | Note, the ZIP log-likelihood is not directly comparable.              |
 | ZIP model with nonzero Q does not encompass the others.               |
 | Vuong statistic for testing ZIP vs. unaltered model is      4.1446    |
 | Distributed as standard normal. A value greater than                  |
 | +1.96 favors the zero altered Z.I.Neg_Bin model.                      |
 | A value less than -1.96 rejects the ZIP model.                        |
 +-----------------------------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Poisson/Negbin regression model				
Constant	-1.132160907	.27064664	-4.183	.0000	
SR520	-.2347356955	.77650707E-01	-3.023	.0025	.13725490
X14	-.3953279550E-01	.51871475E-01	-.762	.4460	.62254902
X15	-.2433312602E-02	.45159412E-02	-.539	.5900	7.1519608
X17	.8299799896	.18034207	4.602	.0000	1.6107843
Alpha	.1575228125	.12823113	1.228	.2193	
	Zero inflation model				
Tau	-15.58052825	7.7538129	-2.009	.0445	

```
--> poisson;lhs=x5
    ;rhs=one,sr520,x14,x15,x17
    ;rh2=one,sov
    ;zip=normal
    ;limit=6;truncation;upper$
```

Zero Altered Poisson Regression Model
Normal distribution used for splitting model.
ZAP term in probability is F[tau x Z(i)]
Comparison of estimated models

	Pr[0|means]	Number of zeros		Log-likelihood
Poisson	.56606	Act.= 126	Prd.= 115.5	-226.14085
Z.I.Poisson	.57118	Act.= 126	Prd.= 116.5	-220.81374

Note, the ZIP log-likelihood is not directly comparable.
ZIP model with nonzero Q does not encompass the others.
Vuong statistic for testing ZIP vs. unaltered model is 1.8530
Distributed as standard normal. A value greater than
+1.96 favors the zero altered Z.I.Poisson model.
A value less than -1.96 rejects the ZIP model.

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Poisson/Negbin regression model				
Constant	-3.492857740	.41471885	-8.422	.0000	
SR520	-.3855394981	.22534217	-1.711	.0871	.13725490
X14	-.2250906487	.14615471	-1.540	.1235	.62254902
X15	.3696268826E-01	.17421499E-01	2.122	.0339	7.1519608
X17	1.912795283	.20605998	9.283	.0000	1.6107843
	Zero inflation model				
Constant	-.2357815939	.28973879	-.814	.4158	
SOV	-1.295511133	.71119258	-1.822	.0685	.59313725