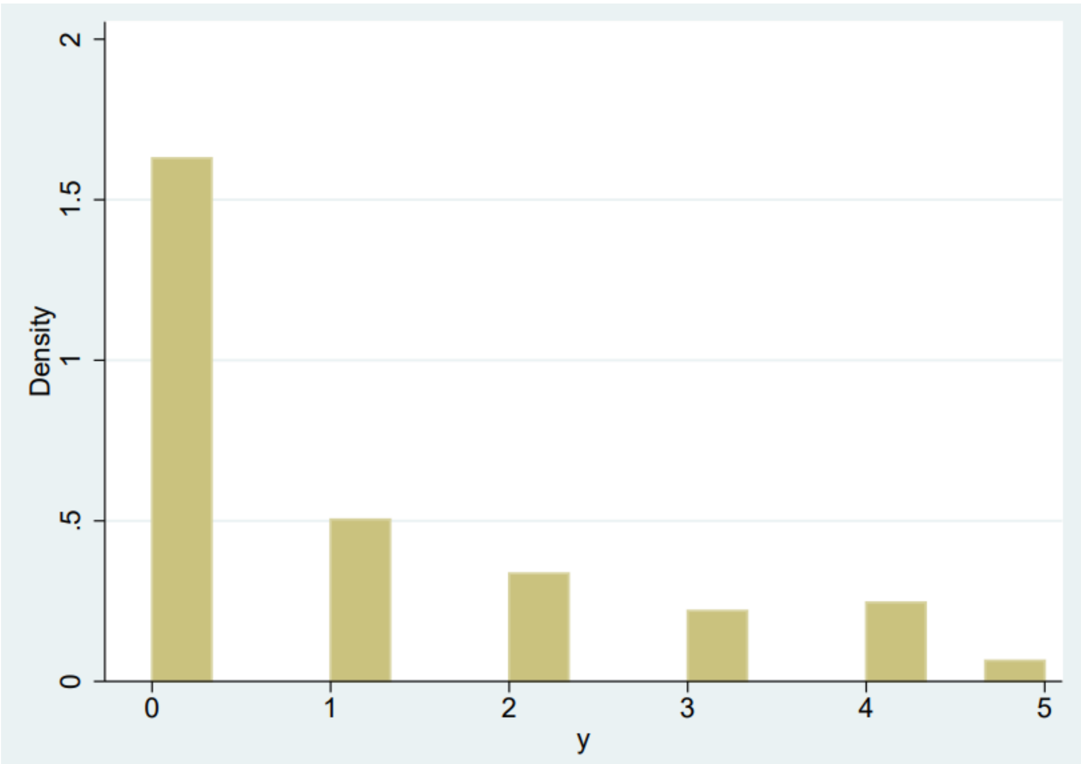
6204641218 Pornwarat Foongpiriya

1 Estimate models for y_i assuming that the model is traditional linear regression model. Interpret your estimated result.

```
. reg y x1 x2 x3 x4

      Source |       SS           df       MS      Number of obs   =       232
-------------+----------------------------------   F(4, 227)       =      5.96
       Model |  44.7298499         4  11.1824625   Prob > F        =    0.0001
    Residual |  425.748598       227  1.87554449   R-squared       =    0.0951
-------------+----------------------------------   Adj R-squared   =    0.0791
       Total |  470.478448       231  2.03670324   Root MSE        =    1.3695

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1016201   .0435073     2.34   0.020     .0158904    .1873499
          x2 |   .1345044   .0462142     2.91   0.004     .0434407     .225568
          x3 |  -.0748194   .0480457    -1.56   0.121    -.1694919    .0198531
          x4 |   .1684563   .0688243     2.45   0.015     .0328401    .3040725
       _cons |   .9568064    .107007     8.94   0.000     .7459523     1.16766
------------------------------------------------------------------------------

. est store ols
```

Ans. The signs of all coefficients are positive as well as they are significant under 95% confidence level except for x3. The overall test is significant. R2 and adjusted-R2 are quite low.

```
. histogram y
(bin=15, start=0, width=.33333333)
```



Ans. From histogram, it can be seen that there is an excess zero problem.

3 Estimate models for y_i assuming that the probability functions follow Poisson probability distribution. Perform GOF test and determine whether Poisson is appropriated in this case. Interpret the estimated result (sign and meaning (in term of incidence-rate ratios), overall test, individual test, pseudo R^2, marginal effects).

```
. poisson y x1 x2 x3 x4, nolog

Poisson regression                              Number of obs     =        232
                                                LR chi2(4)        =      43.33
                                                Prob > chi2       =     0.0000
Log likelihood = -342.88107                     Pseudo R2         =     0.0594
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.0971474	.0306762	3.17	0.002	.0370231	.1572717
x2	.1293024	.0330916	3.91	0.000	.064444	.1941607
x3	-.0715533	.0342177	-2.09	0.037	-.1386187	-.0044879
x4	.1734482	.0507707	3.42	0.001	.0739395	.2729569
_cons	-.1284876	.0849064	-1.51	0.130	-.294901	.0379259

```
. est store poisson

. estat gof

         Deviance goodness-of-fit =  409.4921
         Prob > chi2(227)         =    0.0000

         Pearson goodness-of-fit  =  423.3541
         Prob > chi2(227)         =    0.0000
. poisson y x1 x2 x3 x4, ir nolog

Poisson regression                              Number of obs     =        232
                                                LR chi2(4)        =      43.33
                                                Prob > chi2       =     0.0000
Log likelihood = -342.88107                     Pseudo R2         =     0.0594
```

y	IRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	1.102023	.0338059	3.17	0.002	1.037717	1.170314
x2	1.138034	.0376594	3.91	0.000	1.066566	1.214291
x3	.9309467	.0318548	-2.09	0.037	.8705599	.9955222
x4	1.189399	.0603866	3.42	0.001	1.076742	1.313844
_cons	.8794245	.0746687	-1.51	0.130	.7446053	1.038654

```
. mfx

Marginal effects after poisson
      y  = Predicted number of events (predict)
         =  .95621703
```

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.092894	.02882	3.22	0.001	.036403	.149385	-.317697
x2	.1236411	.0308	4.01	0.000	.063277	.184005	.812709
x3	-.0684205	.03246	-2.11	0.035	-.13204	-.004801	-.818103
x4	.1658541	.04736	3.50	0.000	.073022	.258686	-.28275

Ans. After performing GOF test, the null hypothesis is rejected. As a result, the model is not Poisson distribution. For the estimated result, signs of all coefficients are positive except for x3. They are significant under 95% confidence level. The overall test is significant. Pseudo R2 is 0.0594. All incidence-rate ratios, except that of X3, are greater than one in which each Xs are positively related to Y – negative relationship for X3. Looking at marginal effect, x1 x2 x4 also have positive impact on y where x3 has the opposite way.

4 Estimate models for y_i assuming that the probability functions follow Negative Binomial probability distribution. Determine whether Negative Binomial regression model is appropriated in this case. Interpret your estimated result (sign and meaning (in term of incidence-rate ratios), overall test, individual test, pseudo R^2, marginal effects).

```
. nbreg y x1 x2 x3 x4, nolog

Negative binomial regression                    Number of obs     =        232
                                                LR chi2(4)        =      21.24
Dispersion     = mean                           Prob > chi2       =     0.0003
Log likelihood = -317.49278                     Pseudo R2         =     0.0324
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.1285534	.0506934	2.54	0.011	.0291962	.2279106
x2	.151011	.0506477	2.98	0.003	.0517434	.2502785
x3	-.0672859	.0481376	-1.40	0.162	-.1616339	.0270621
x4	.1726312	.0707035	2.44	0.015	.034055	.3112075
_cons	-.1435596	.1177204	-1.22	0.223	-.3742874	.0871682
/lnalpha	.0479945	.2389531			-.4203449	.5163339
alpha	1.049165	.2507012			.6568202	1.675872

```
Likelihood-ratio test of alpha=0:  chibar2(01) =   50.78 Prob>=chibar2 = 0.000

. est store nb

. nbreg y x1 x2 x3 x4, ir nolog

Negative binomial regression                    Number of obs     =        232
                                                LR chi2(4)        =      21.24
Dispersion     = mean                           Prob > chi2       =     0.0003
Log likelihood = -317.49278                     Pseudo R2         =     0.0324
```

y	IRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	1.137182	.0576476	2.54	0.011	1.029627	1.255973
x2	1.163009	.0589037	2.98	0.003	1.053105	1.284383
x3	.9349279	.0450052	-1.40	0.162	.8507526	1.027432
x4	1.188428	.084026	2.44	0.015	1.034641	1.365072
_cons	.8662692	.1019776	-1.22	0.223	.6877792	1.09108
/lnalpha	.0479945	.2389531			-.4203449	.5163339
alpha	1.049165	.2507012			.6568202	1.675872

```
Likelihood-ratio test of alpha=0:  chibar2(01) =   50.78 Prob>=chibar2 = 0.000

. mfx

Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =  .94607122
```

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.1216207	.04796	2.54	0.011	.02763	.215611	-.317697
x2	.1428671	.04796	2.98	0.003	.048862	.236872	.812709
x3	-.0636573	.04556	-1.40	0.162	-.152956	.025641	-.818103
x4	.1633214	.06686	2.44	0.015	.032269	.294374	-.28275

Ans. From the test of alpha=0. The null hypothesis is rejected. As a result, negative binomial regression model should be applied in this case. The signs of all coefficients are positive as well as they are significant under 95% confidence level except for x3. The overall test is significant. Pseudo R2 is 0.0324. All incidence-rate ratios, except that of X3, are greater than one in which each Xs are positively related to Y – negative relationship for X3. Looking at marginal effect, x1 x2 x4 also have positive impact on y where x3 has the opposite way.

5 Estimate models for y_i assuming that the model is Zero Inflated Poisson (x_{1i}, x_{2i}, and x_{3i} are independent variables in Poisson model and x_{4i} is independent variable in Inflated (Logit) model). Interpret your estimated result. Determine which model (Linear regression model, Poisson, Negative Binomial, or ZIP) is the most appropriated model in this case? Why? (provide the tests)

```
. zip y x1 x2 x3, inflate(x4) vuong

Fitting constant-only model:

Iteration 0:   log likelihood = -355.48956
Iteration 1:   log likelihood =  -321.8304
Iteration 2:   log likelihood = -317.80147
Iteration 3:   log likelihood = -317.79274
Iteration 4:   log likelihood = -317.79274

Fitting full model:

Iteration 0:   log likelihood = -317.79274
Iteration 1:   log likelihood = -312.73621
Iteration 2:   log likelihood =  -312.6159
Iteration 3:   log likelihood =  -312.6158
Iteration 4:   log likelihood =  -312.6158

Zero-inflated Poisson regression                 Number of obs     =       232
                                                 Nonzero obs       =       106
                                                 Zero obs          =       126

Inflation model = logit                          LR chi2(3)        =     10.35
Log likelihood  = -312.6158                      Prob > chi2       =    0.0158
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
y						
x1	.0805446	.0398159	2.02	0.043	.0025068	.1585824
x2	.0857883	.0372107	2.31	0.021	.0128567	.1587199
x3	-.0672468	.0357098	-1.88	0.060	-.1372367	.002743
_cons	.4589728	.1106031	4.15	0.000	.2421947	.6757508
inflate						
x4	-.2738532	.1212311	-2.26	0.024	-.5114618	-.0362446
_cons	-.3379298	.1908217	-1.77	0.077	-.7119334	.0360738

```
Vuong test of zip vs. standard Poisson:            z =     3.92  Pr>z = 0.0000
. mfx

Marginal effects after zip
      y  = Predicted number of events (predict)
         =   .9868501
```

variable	dy/dx	Std. Err.	z	P>|z|	[95%	C.I.]	X
x1	.0794855	.03893	2.04	0.041	.003192	.155779	-.317697
x2	.0846602	.03606	2.35	0.019	.013992	.155328	.812709
x3	-.0663626	.03507	-1.89	0.058	-.135102	.002377	-.818103
x4	.1176249	.05326	2.21	0.027	.013231	.222019	-.28275

Ans. According to Vuong test, the null hypothesis is rejected. As a result, zero-inflated poisson model should be applied rather than poisson. The signs of all coefficients are positive as well as they are significant under 95% confidence level except for x3. However, they all are significant under 90% confidence level. The overall test is significant. The loglikelihood value is -312.6158. Looking at marginal effect, x1 x2 x4 have positive impact on Y where x3 has the opposite way.

Which model is the most appropriated?

From the results :

1. GOF => the model is not Poisson
2. Overdispersion test => Negative Binomial Model should be applied rather than Poisson
3. Voung test => Zero-inflated Poisson Model should be applied rather than Poisson

As the choice of Poisson can be eliminated, we compare NB model and ZIP model. Looking at the significance of each independent variables, they are all significant in ZIP model. Moreover, from the histogram it can be seen that the data can be considered as not dispersed where the problem of high variance might not exist, and there is excess zero. As a result, I think, ZIP should be used in this case.