Assignment 11
Truncated Regression & Tobit Models

The model

According to the following model,

$$y_{ki} = \beta_{k0} + \beta_{k1} x_i + u_{ki} \quad (1)$$

where: y_{ki} is Dependent variable, k=1, 2, 3.
x_i is Independent variable
u_i is Stochastic disturbance term

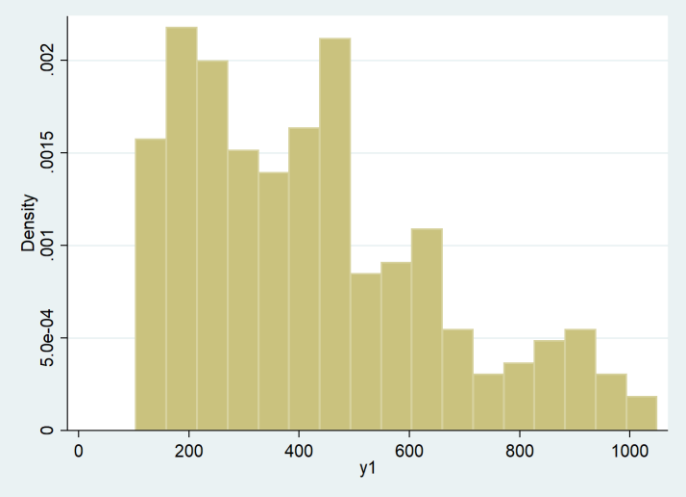
Requirements: (assign11.dta)

1 Plot histogram of y_{1i}, y_{2i}, y_{3i}, compute descriptive statistics of these three variables, then determine limitation of these three dependent variables.

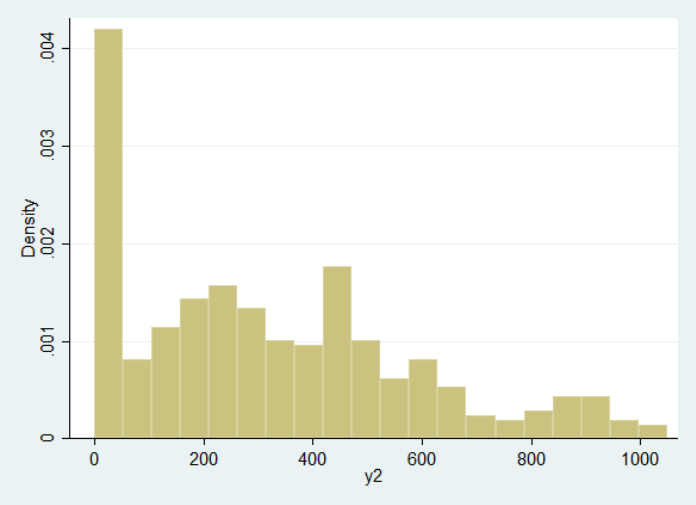
```
. sum y1 y2 y3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          y1 |        297    423.1683    227.8976   103.4663   1049.314
          y2 |        400    317.5914    266.3419          0   1049.314
          y3 |        400     10446.1    23186.57  -866.3348   99508.07
```

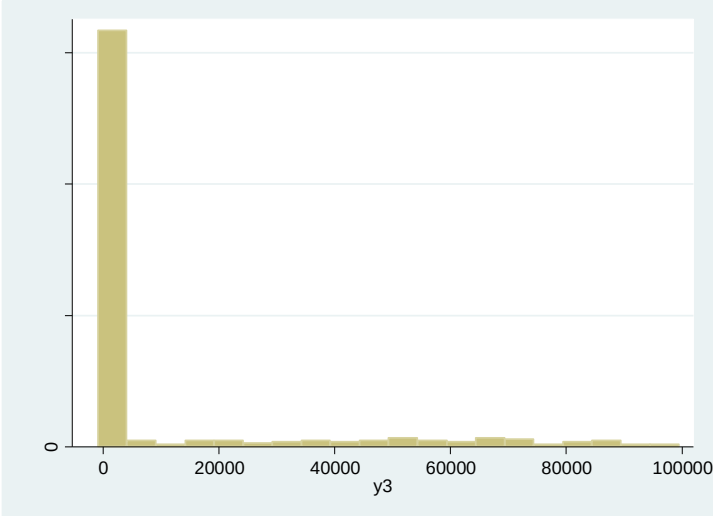
```
. histogram y1
```



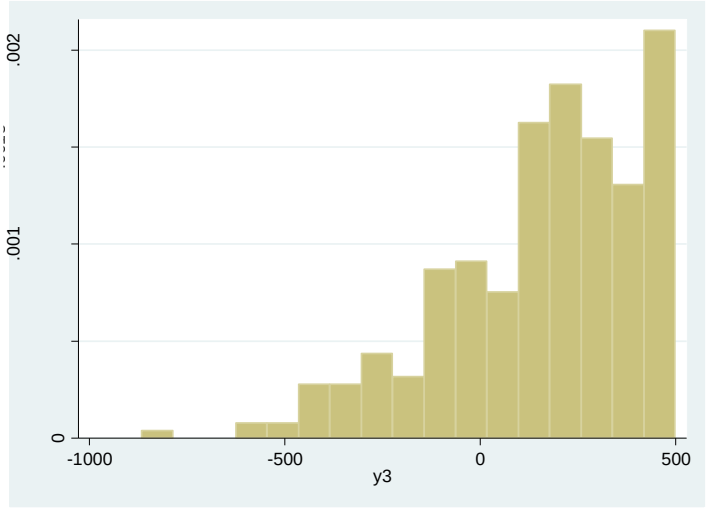
```
. histogram y2
```




```
. histogram y3
```




```
.  histogram y3 if y3<1000
```



According to the summarized statistics table and histograms, it might be the case that there exists the truncated problem for y1 (the value of y below 103 cannot be observed). For y2, we can clearly observe that the dependent variable can be fully observed but the value equal to zero has significantly high density. Therefore, this can be the case that y2 has a censored data problem. For y3, it is highly likely to be the case of outlier problem.

2 Estimate the model (1) for y_{1i}, y_{2i}, y_{3i} using OLS, using truncated regression model, Tobit model, determine the most appropriated models for each y_{ki}.

***y1 ols**

```
. reg y1 x
      Source |       SS           df       MS      Number of obs   =       297
-------------+----------------------------------   F(1, 295)       =     30.30
       Model |  1431905.59         1  1431905.59   Prob > F        =    0.0000
    Residual |  13941534.7       295  47259.4396   R-squared       =    0.0931
-------------+----------------------------------   Adj R-squared   =    0.0901
       Total |  15373440.3       296  51937.2982   Root MSE        =    217.39

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   123.1693   22.37637     5.50   0.000     79.13176    167.2069
       _cons |   69.01677   65.56422     1.05   0.293    -60.01611    198.0496
------------------------------------------------------------------------------
. est store y1_ols
```

***y1 trunc**

```
. truncreg y1 x, ll(103)
(note: 0 obs. truncated)


Fitting full model:
Iteration 0:   log likelihood = -1993.4742
. . .
Iteration 4:   log likelihood = -1973.9423

Truncated regression
Limit:   lower =        103                     Number of obs     =        297
         upper =       +inf                     Wald chi2(1)      =      23.38
Log likelihood = -1973.9423                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
```

```
          y1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   249.6062   51.61679     4.84   0.000     148.4392    350.7733
       _cons |  -456.6913   178.5338    -2.56   0.011    -806.6112   -106.7714
-------------+----------------------------------------------------------------
      /sigma |   306.4488   26.46552    11.58   0.000     254.5774    358.3203
------------------------------------------------------------------------------
```

```
. est store y1_trunc
```

***Test whether truncated model should be employed**

```
. lrtest y1_ols y1_trunc, force

Likelihood-ratio test                                 LR chi2(1)  =     89.69
(Assumption: y1_ols nested in y1_trunc)               Prob > chi2 =    0.0000
```

According to the LR-test, the truncated model is more appropriated for y1.

***y2 ols**

```
. reg y2 x
      Source |       SS           df       MS      Number of obs   =       400
-------------+----------------------------------   F(1, 398)       =     76.58
       Model |  4567087.34         1  4567087.34   Prob > F        =    0.0000
    Residual |  23737174.2       398  59641.1413   R-squared       =    0.1614
-------------+----------------------------------   Adj R-squared   =    0.1592
       Total |  28304261.6       399  70937.9989   Root MSE        =    244.22

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   179.8325   20.55046     8.75   0.000     139.4315    220.2335
       _cons |  -179.3679   58.08821    -3.09   0.002     -293.566    -65.1698
------------------------------------------------------------------------------
```

```
. est store y2_ols
```

***y2 Tobit**

```
. tobit y2 x, ll(0)

Tobit regression                                  Number of obs     =        400
                                                  LR chi2(1)        =      73.17
                                                  Prob > chi2       =     0.0000
Log likelihood = -2389.3077                       Pseudo R2         =     0.0151

------------------------------------------------------------------------------
          y2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   219.1562   24.86468     8.81   0.000     170.2741    268.0383
       _cons |  -318.6265   70.93087    -4.49   0.000    -458.0714   -179.1816
-------------+----------------------------------------------------------------
      /sigma |   284.9073   11.56717                      262.1671    307.6475
------------------------------------------------------------------------------

            72  left-censored observations at y2 <= 0
           328     uncensored observations
             0 right-censored observations

. est store y2_tobit
```

***Test whether tobit model should be employed**

```
. lrtest y2_ols y2_tobit, force

Likelihood-ratio test                                 LR chi2(1)  =    752.97
(Assumption: y2_ols nested in y2_tobit)               Prob > chi2 =    0.0000
```

According to the LR-test, the tobit model is more appropriated for y2.

. *y3 ols

```
. reg y3 x

      Source |       SS           df       MS      Number of obs   =       400
-------------+----------------------------------   F(1, 398)       =     26.43
       Model |  1.3359e+10         1  1.3359e+10   Prob > F        =    0.0000
    Residual |  2.0115e+11       398   505402183   R-squared       =    0.0623
-------------+----------------------------------   Adj R-squared   =    0.0599
```

```
       Total |  2.1451e+11       399   537616802   Root MSE        =     22481

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   9726.042   1891.765     5.14   0.000     6006.941    13445.14
       _cons |  -16431.39   5347.288    -3.07   0.002    -26943.85   -5918.932
------------------------------------------------------------------------------
. est store y3_ols

. *y3 tobit
. tobit y3 x, ul(1000)

Tobit regression                                Number of obs     =        400
                                                LR chi2(1)        =      63.89
                                                Prob > chi2       =     0.0000
Log likelihood = -2469.6369                     Pseudo R2         =     0.0128

------------------------------------------------------------------------------
          y3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |         334   40.73749     8.20   0.000     253.9131    414.0869
       _cons |   -530.425    114.018    -4.65   0.000     -754.576    -306.274
-------------+----------------------------------------------------------------
      /sigma |   467.1217   19.91449                      427.9713    506.2722
------------------------------------------------------------------------------
             0  left-censored observations
           314     uncensored observations
            86 right-censored observations at y3 >= 1000
. est store y3_tobit
. *Test whether tobit model should be employed
. lrtest y3_ols y3_tobit, force

Likelihood-ratio test                                 LR chi2(1)  =   4210.22
(Assumption: y3_ols nested in y3_tobit)               Prob > chi2 =    0.0000
```

According to the LR-test, the tobit model is more appropriated for y3.