Faculty of Economics and Political science
Cairo University

Garbage Disposal: A Case Study on Ismailia Governorate's citizens

Name:
Ghofran Talaat Sayed Mohameed

Under supervision of:

Dr. Sally Abdelfatah

Jan-2022

Table of contents

Contents

Abstract

Proper means of garbage disposal is considered one of the most important things in our life and that is for many reasons although some people do not even notice that. However, neglecting it will have negative effects on our life. In this paper we are mainly concerned with how can the person's education, work status and activity can be related to or affecting the means of garbage disposal. The most important results we have obtained is that there is a significant relation between proper means of garbage disposal and the person's work status as well as between proper means of garbage disposal and the person's education but we found that there is no significant relationship between proper means of garbage disposal and the person's activity.

In this paper we are going to fit a binary logistic model, test the goodness of fit of our model, test the model significance, determine the predictive power of our model, make a two-way contingency tables, three-way contingency table, conduct inference in order to test our hypotheses of independence/ association and finally conducting some variable's description.

Section (1)

Introduction:

Garbage is classified by source and composition. Generally speaking, waste materials are either liquid or solid in form, and their components may be either hazardous (i.e. medical waste) or inert in their effects on health and the environment. Proper garbage disposal involves committing ourselves to make a waste segregation; sort out materials that can be re-used and recycled before dumping the old things and scraps to the bin and for the leftover waste it can be then sorted into organic waste, inorganic waste, non-recyclable material, and even compost that can be packaged and sold. There are many reasons why we have to properly dispose of garbage like promoting health and protecting the environment.

We are going to mention only two examples on how poor garbage management can have a serious impact on both the environment and our health.

1.1 Garbage and environment:

Ineffective garbage management has a lot to do with climate change and air pollution. Also, it can directly affect the larger ecosystem and various species. Landfills, is seen as the last resort in the waste hierarchy, as it releases methane; a strong greenhouse gas associated with climate change. Methane is created by microorganisms which can be found in landfills from biodegradable waste, like: food, paper and garden waste. Following how landfills are built, it might also pollute soil and water. Waste is transported and treated after its' collection. Carbon dioxide and air pollutants are released into the atmosphere through the transportation process. Heat or electricity as forms of energy can be generated from waste. This in turn may reduce the energy produced by other fuels. Thus, energy generated from waste can help in reducing emissions of greenhouse gas.

1.2 Medical garbage and health:

According to WHO[1], health-care providers' waste is between 75% and 90% and it is considered as non-risk or "general" health-care waste, in comparison with domestic waste. This is generated mostly from the functions of administrating and housekeeping of health-care premises. Also, it may contain waste generated during preservation of health-care establishments. The other 10–25% of healthcare waste is considered as hazardous and may impose serious health risks.

Health-care waste contains a large amount of general waste and a smaller amount of hazardous waste. People who are exposed to hazardous health-care waste are at a greater risk, especially those within health-care premises that generate hazardous waste. Also, people outside these premises will be exposed to it as a result of incautious management, as they might got

[1] Pruss-Ustun, A., Giroult, E., Rushbrook, P., & World Health Organization. (1999). Safe management of wastes from health-care activities. World Health Organization.

infected due to the exposure to health-care wastes. So, it is important to illustrate the proper means of garbage disposal.

In this paper, we are going to analyze data of the Income, Expenditure and Consumption Survey in Ismailia governorate, Egypt for year 1995/1996 through studying the variables that are related to our research question that will be mentioned later on. Before we start anything we have to mention the different variables we are going to deal with in our study.

1.3 Variables overview:

Variable Name	Scale of Measurement	Base/ Reference Category in the binary logistic model
Explanatory variables		
work status of person	Nominal: 1= "waged job" 2= "self-employed employing others" 3= "self-employed not employing others"	"waged job"
education of person	Ordinal: 2= "illiterate" 3= "read and write" 4= "basic education" 5= "secondary education" 6= "diploma" 7= "university degree" 8= "higher than university degree"	"illiterate"
activity of person	Nominal: 1= "agriculture" 3= "manufacturing" 4= "electricity" 5= "construction" 6= "trade" 7= "transportation" 8= "finance" 9= "services"	
Response variable		
proper means of garbage disposal	Binary: 0 ="Yes" 1 = "No"	

In this paper, we are studying the different existing relations and connections between "proper means of garbage disposal" and others explanatory variables that are assumed to affect "proper means of garbage disposal" according to our research questions.

1.4 Research questions:

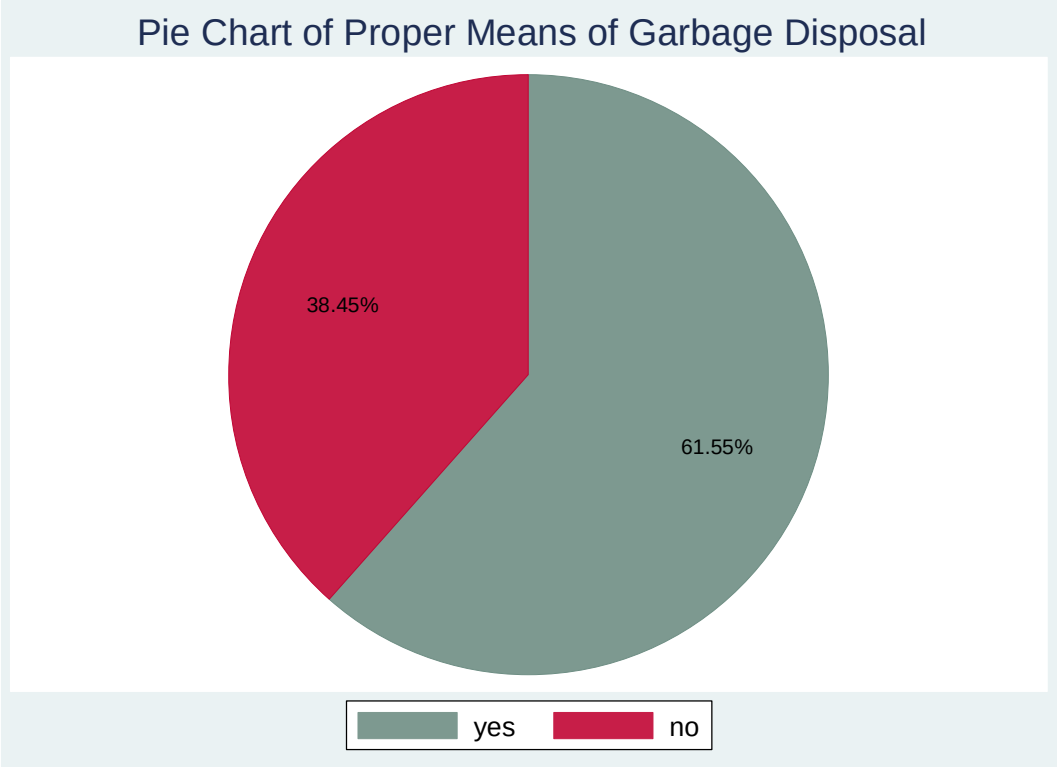
- Are more educated people tending to use the proper means of garbage disposal?
- Is the person's activity has an impact on its behavior regarding garbage disposal?
- Can we consider a relationship between person's work status and his attitude regarding garbage disposal?
- Can we look at the impact of the person's work status on garbage disposal taking into consideration the person's education?
- Can we fit a binary logistic model? Will it be significant?
- Will our model have a good fit?
- Will our model have a good predictive power?

Section (2)

Variables Description:

2.1 Response Variable (Proper means of garbage disposal):

Originally this variable took values 0 for "No"; those who don't use the proper means of garbage disposal, 1 for "Yes"; those who use the proper means of garbage disposal. But we have reversed the order to be 0 for "Yes" and 1 for "No" and the reason behind this reversal will be explained later on.

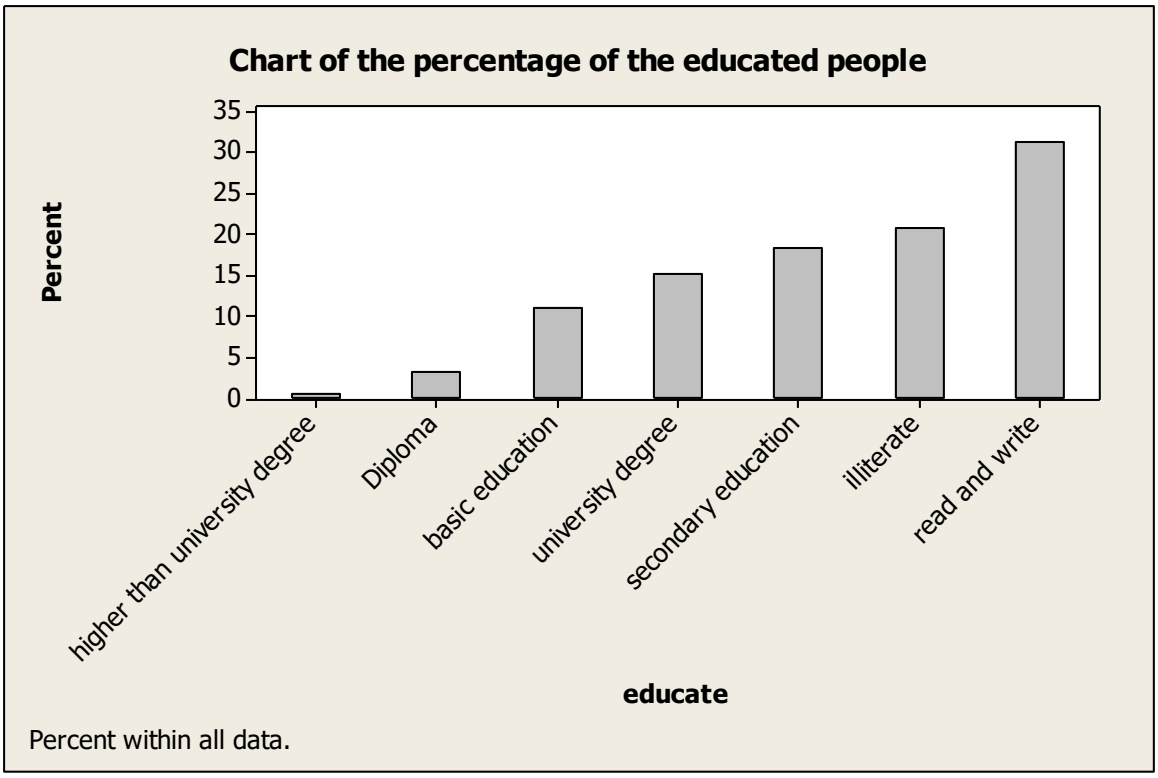


As we can see from the "Pie Chart" we can find that the largest percentage is for those who use the proper means of garbage disposal (61.55%).

2.2 Explanatory variables:

2.2.1 Education of person:

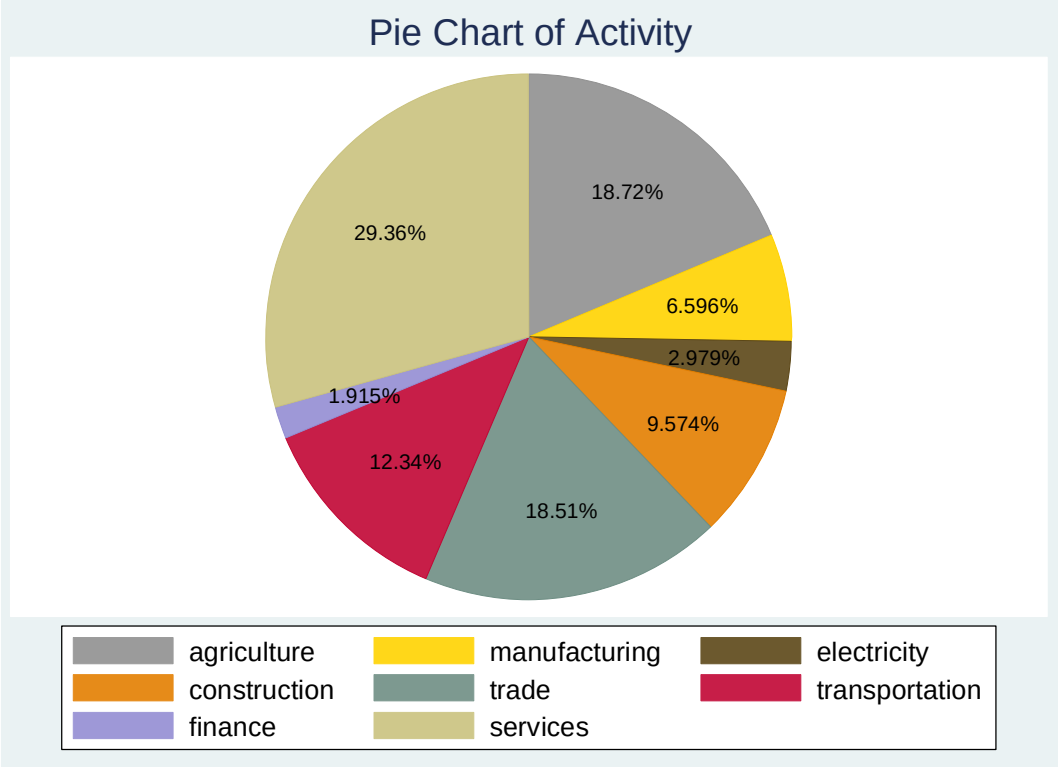
This variable takes values 2 for illiterate people, 3 for those who read and write only, 4 for those who have a basic education, 5 for those who have a secondary education, 6 for those who had a Diploma, 7 for those who had a university degree, and 8 for those who got a higher than university degree.



As we can see from the "Bar Chart", in our data the highest percentage of people goes for those who read and write (approximately 33%) and the lowest one is for those who got a higher than university degree (approximately 1%).

2.2.2 Activity of person:

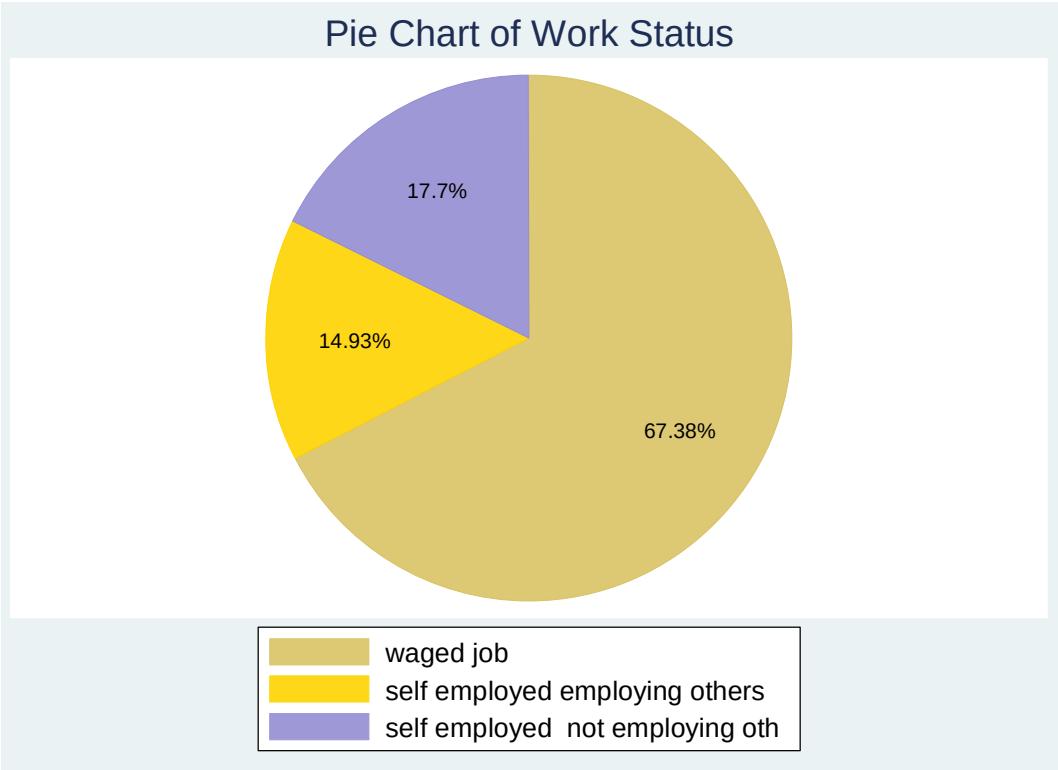
This variable takes values 1 for those whose activity is related to agriculture, 3 for those whose activity is related to manufacturing, 4 for those whose activity is related to electricity, 5 for those whose activity is related to construction, 6 for those whose activity is related to trade, 7 for those whose activity is related to transportation, 8 for those whose activity is related to finance, and 9 for those whose activity is related to services.



As we can see from the “Pie Chart”, the largest percentage of people are those who have an activity related to services (29.36%), and the smallest percentage of people are those who have an activity related to finance (1.915%).

2.2.3 Work status of person:

This variable takes values 1 for those who have a waged job, 2 for those who are self-employed and employing others, and 3 for those who are self-employed but not employing others.



As we can see from the "Pie Chart", the largest percentage of people have a waged job (67.38%) but the smallest percentage of people's work status goes for those who are self-employed and employing others (14.93%).

Section (3)

Inferences:

In this part we are going to make Inference for Two-way and Three-way contingency tables respectively.

3.1 Inference for Two-way contingency tables:

In this part we are going to study the relationship between our explanatory variables and the response. But before starting our analysis we found that the order of the response variable's categories was inappropriate for our analysis and our interpretation that we are going to make later on as a result we reversed the order.

It was as follows:

```
. tab garbage

    Proper
  means of
   garbage
  disposal |      Freq.     Percent        Cum.
-----------+-----------------------------------
        no |        203       38.45       38.45
       yes |        325       61.55      100.00
-----------+-----------------------------------
     Total |        528      100.00
```

It became as follows:

```
. tab Garbage

   Garbage |      Freq.     Percent        Cum.
-----------+-----------------------------------
       yes |        325       61.55       61.55
        no |        203       38.45      100.00
-----------+-----------------------------------
     Total |        528      100.00
```

Although both are appropriate to work with them but we are always seeking a simpler interpretation.

<u>Analyzing our data:</u>

For our three explanatory variables we will do the following:

- Two-way contingency table: We know that contingency means concurrency of events, hence the two-way contingency table between Proper means of garbage disposal and the different explanatory variables will help us to know by how much do our different events tend to occur together but this table will be provided in the appendix part.
- Test of independence.
- Test of association (depending on the explanatory type; Nominal or Ordinal)

3.1.1 proper means of garbage disposal and the person's work status:

3.1.1.1 Test of independence between Proper means of garbage disposal and the person's work status:

Since proper means of garbage disposal is a binary variable and the person's work status is a nominal variable, and we don't have dependent samples hence we worked with "Person's Chi-Squared" Test of Independence as follows

Step(1): Stating the null and the alternative hypotheses

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j} \quad \forall\, I\&J$$

$$H_1: \pi_{ij} \neq \pi_{i+}\pi_{+j} \quad for\ at\ least\ one\ combination\ of\ I\ \&\ J$$

Step (2): STATA output:

```
. tab wrkstat Garbage, chi2

                        |          Garbage
Work Status of Person   |       yes          no |     Total
------------------------+----------------------+----------
            waged job   |       199         117 |       316
self employed employi   |        30          40 |        70
self employed  not em   |        51          32 |        83
------------------------+----------------------+----------
                  Total |       280         189 |       469

          Pearson chi2(2) =   9.7675   Pr = 0.008
```

Step (3): Comment

Since p-value is equal to 0.008 which is less than the significant level $\propto$ "0.05" we are going to reject the null hypothesis at 95% confidence. We are 95% confidant that there is a significant relation between X and Y (Proper means of garbage disposal "Y" and the person's work status "X"). We know that any test of independence says nothing about the nature of the

association if it exists hence it is important to investigate the nature of the association using the appropriate measures of association and we are going to do that through Cramer's V since "X" is nominal variable.

3.1.1.2 Test of association between Proper means of garbage disposal and the person's work status:

Based on STATA we found the following,

```
. tab wrkstat Garbage, V

                        |          Garbage
Work Status of Person   |       yes          no |     Total
------------------------+----------------------+----------
            waged job   |       199         117 |       316
self employed employi   |        30          40 |        70
self employed  not em   |        51          32 |        83
------------------------+----------------------+----------
                Total   |       280         189 |       469

              Cramér's V =   0.1443
```

Comment:

Since $Cramer's\ V = 0.1443$ hence, we can say that there is a weak relationship between X and Y (Proper means of garbage disposal "Y" and the person's work status "X").

3.1.2 Proper means of garbage disposal and the person's activity:

3.1.2.1 Test of independence between Proper means of garbage disposal and the person's activity:

Since Proper means of garbage disposal is a binary variable and the person's activity is a nominal variable, and we don't have a dependent samples hence we worked with Person's Chi-Squared Test of Independence as follows,

Step(1): Stating the null and the alternative hypotheses

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j} \quad \forall\ I\&J$$

$$H_1: \pi_{ij} \neq \pi_{i+}\pi_{+j} \quad for\ at\ least\ one\ combination\ of\ I\ \&\ J$$

Step (2): STATA output:

```
. tab activity Garbage, chi2
```

Activity of Person	Garbage yes	no	Total
agriculture	42	46	88
manufacturing	21	10	31
electricity	9	5	14
construction	27	18	45
trade	54	33	87
transportation	35	23	58
finance	8	1	9
services	84	54	138
Total	280	190	470

Pearson chi2(7) = 9.6662 Pr = 0.208

Step (3): Comment

Since p-value is equal to 0.208 which is greater than the significant level ∝ "0.05". Hence at level of confidence 95% we don't have enough evidence to say that there is a significant relationship between X and Y (Proper means of garbage disposal "Y" and the person's activity "X"), however it is more preferable to test for association.

3.1.2.2 Test of association between Proper means of garbage disposal and the person's activity:

Based on STATA we found the following,

```
. tab activity Garbage, V
```

Activity of Person	Garbage yes	no	Total
agriculture	42	46	88
manufacturing	21	10	31
electricity	9	5	14
construction	27	18	45
trade	54	33	87
transportation	35	23	58
finance	8	1	9
services	84	54	138
Total	280	190	470

Cramér's V = 0.1434

Comment:

Since $Cramer's\ V = 0.1434$ hence, we can say that there is a very weak relationship between X and Y (Proper means of garbage disposal "Y" and the person's activity "X").

3.1.3 Proper means of garbage disposal and the person's education:

3.1.3.1 Test of independence between Proper means of garbage disposal and the person's activity:

Since Proper means of garbage disposal is a binary variable and the person's education is an ordinal variable, and we don't have a dependent samples hence we worked with Person's Chi-Squared Test of Independence as follows,

Step(1): Stating the null and the alternative hypotheses

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j} \quad \forall\, I\&J$$

$$H_1: \pi_{ij} \neq \pi_{i+}\pi_{+j} \quad for\ at\ least\ one\ combination\ of\ I\ \&\ J$$

Step (2): STATA output:

Education of Person	Garbage yes	no	Total
illiterate	53	57	110
read and write	86	79	165
basic education	40	18	58
secondary education	60	36	96
Diploma	15	2	17
university degree	70	10	80
higher than universit	1	1	2
Total	325	203	528

Pearson chi2(6) = 43.8818 Pr = 0.000

Step (3) Comment:

Since p-value is equal to 0.000 which is less than the significant level $\propto$ "0.05" we are going to reject the null hypothesis at 95% confidence. We are 95% confidant that there is a significant relation between X and Y (Proper means of garbage disposal "Y" and the person's education "X"). We are going investigate the nature of the association using the appropriate

measures of association which is in this case Goodman and Kruskal Gamma since "X" is ordinal variable.

3.1.3.2 Test of association between proper means of garbage disposal and the person's education:

Based on STATA we found the following,

```
. tab educate Garbage, gamma
```

Education of Person	Garbage yes	no	Total
illiterate	53	57	110
read and write	86	79	165
basic education	40	18	58
secondary education	60	36	96
Diploma	15	2	17
university degree	70	10	80
higher than universit	1	1	2
Total	325	203	528

gamma = -0.3730 ASE = 0.057

Comment:

Since $\gamma = -0.3730$ hence, we can say that there is a moderate positive relationship between using the proper means of garbage disposal and the person's education, hence more educated individuals tend to dispose garbage with proper means.

3.2 Three-way contingency table:

We made a three-way contingency table between work status, Garbage, and education variables in order to check if education will affect the relation between Garbage and work status. But we found that for each category of the variable "education" we have 3×2 partial table, hence we were not able to apply any of the methods that we have learned, as a result we have decided to study this relation between the three variables through fitting a binary logistic regression model and this will have a separate section for it.

```
. table wrkstat Garbage, by(educate)
```

Education of Person and Work Status of Person	Garbage yes	Garbage no
illiterate		
waged job	20	21
self employed employing others	9	20
self employed not employing others	17	9
read and write		
waged job	48	48
self employed employing others	10	11
self employed not employing others	18	16
basic education		
waged job	21	14
self employed employing others	1	2
self employed not employing others	6	1
secondary education		
waged job	43	23
self employed employing others	5	6
self employed not employing others	6	6
Diploma		
waged job	11	2
self employed employing others		
self employed not employing others	3	
university degree		
waged job	55	9
self employed employing others	5	
self employed not employing others	1	
higher than university degree		
waged job	1	
self employed employing others		1
self employed not employing others		

Section (4)

Modelling:

4.1 Model choosing

Since our response variable "proper means of garbage disposal" is a binary variable we are going to fit a binary logistic model using STATA.

In order to determine which of our three explanatory variables to be included in our model we used the stepwise procedure (backward stepwise) and the results were as follows:

```
. xi : stepwise, pr(0.05) : logit Garbage i.educate i.activity i.wrkstat
i.educate         _Ieducate_2-8       (naturally coded; _Ieducate_2 omitted)
i.activity        _Iactivity_1-9      (naturally coded; _Iactivity_1 omitted)
i.wrkstat         _Iwrkstat_1-3       (naturally coded; _Iwrkstat_1 omitted)
                      begin with full model
p = 0.8993 >= 0.0500  removing _Ieducate_3
p = 0.8896 >= 0.0500  removing _Iactivity_4
p = 0.9000 >= 0.0500  removing _Iwrkstat_3
p = 0.6674 >= 0.0500  removing _Ieducate_8
p = 0.5839 >= 0.0500  removing _Iactivity_9
p = 0.3379 >= 0.0500  removing _Iactivity_8
p = 0.3083 >= 0.0500  removing _Iactivity_7
p = 0.2823 >= 0.0500  removing _Iactivity_5
p = 0.1514 >= 0.0500  removing _Ieducate_4
p = 0.1866 >= 0.0500  removing _Ieducate_5
p = 0.1248 >= 0.0500  removing _Iactivity_3
p = 0.0951 >= 0.0500  removing _Iactivity_6

Logistic regression                             Number of obs     =        469
                                                LR chi2(3)        =      43.56
                                                Prob > chi2       =     0.0000
Log likelihood = -294.42248                     Pseudo R2         =     0.0689

------------------------------------------------------------------------------
     Garbage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Iwrkstat_2 |    .654064   .2695936     2.43   0.015     .1256703    1.182458
  _Ieducate_7 |  -1.719438   .3728508    -4.61   0.000    -2.450212   -.9886639
  _Ieducate_6 |  -1.693715     .76426    -2.22   0.027    -3.191637   -.1957934
        _cons |  -.2521947   .1125377    -2.24   0.025    -.4727645   -.0316249
------------------------------------------------------------------------------

.
end of do-file
```

Based on the backward stepwise, we are going to include only two variables (work status of person and education of person) out of the three explanatory variables. Hence our model will include 2 independent variables.

As a result if we fitted our binary logistic model will find:

```
. logit Garbage i.educate i.wrkstat, or

Iteration 0:   log likelihood = -316.20142
Iteration 1:   log likelihood = -292.40079
Iteration 2:   log likelihood = -291.93369
Iteration 3:   log likelihood = -291.93167
Iteration 4:   log likelihood = -291.93167

Logistic regression                             Number of obs     =        469
                                                LR chi2(8)        =      48.54
                                                Prob > chi2       =     0.0000
Log likelihood = -291.93167                     Pseudo R2         =     0.0768

------------------------------------------------------------------------------
             Garbage | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             educate |
          read and write |   .9727458   .2611653    -0.10   0.918     .5747336    1.646388
        basic education |   .6063856   .2301909    -1.32   0.188      .288153     1.27607
     secondary education |   .6243805   .1922826    -1.53   0.126     .3414409    1.141782
                 Diploma |   .1488599   .1175852    -2.41   0.016     .0316526    .7000777
       university degree |   .1404734   .0596433    -4.62   0.000     .0611202    .3228521
higher than university degree |   .7650854   1.104286    -0.19   0.853     .0451989    12.95065
                         |
             wrkstat |
self employed employing others |   1.678876   .4801471     1.81   0.070     .9584752    2.940739
self employed  not employing others |   .7475197   .1991549    -1.09   0.275     .4434501    1.260087
                         |
               _cons |   1.008743   .2425132     0.04   0.971     .6297125    1.615917
------------------------------------------------------------------------------
```

4.2 General clarification:

In order to know which category of our variables was used as a reference or a base category we made the following using STATA,

Variable (1) Work Status of person:

```
. tab wrkstat

  Work Status of Person |      Freq.     Percent        Cum.
------------------------+-----------------------------------
                waged job |        316       67.38       67.38
self employed employing others |         70       14.93       82.30
self employed  not employing others |         83       17.70      100.00
------------------------+-----------------------------------
                  Total |        469      100.00
```

Hence, waged job was the base category.

Variable (2) education of person:

```
. tab educate
```

Education of Person	Freq.	Percent	Cum.
illiterate	110	20.83	20.83
read and write	165	31.25	52.08
basic education	58	10.98	63.07
secondary education	96	18.18	81.25
Diploma	17	3.22	84.47
university degree	80	15.15	99.62
higher than university degree	2	0.38	100.00
Total	528	100.00	

Hence, illiterate was the base category.

<u>As a kind of summarization:</u>

Variable Name	Scale of Measurement	Base/ Reference Category
Work Status of person	Nominal: 1= "waged job" 2= "self-employed employing others" 3= "self-employed not employing others"	"waged job"
education of person	Ordinal: 2= "illiterate" 3= "read and write" 4= "basic education" 5= "secondary education" 6= "diploma" 7= "university degree" 8= "higher than university degree"	"illiterate"

Returning to the model interpretation:

Note that: we can interpret $\hat{B}$ *or* $e^{\hat{B}}$, so we are going to interpret the model using $e^{\hat{B}}$ but taking into consideration that the estimated model using STATA can be written as follows:

The estimated model:

$$\pi(x_i) = \frac{e^{\sum_{j=0}^{J} B_j x_{ij}}}{1 - e^{\sum_{j=0}^{J} B_j x_{ij}}}$$

$$\pi(x_i) = \frac{e^{0.0087 - 0.027x_{11} - 0.5x_{12} - 0.47x_{13} - 1.9x_{14} - 1.96x_{15} - 0.27x_{16} + 0.51x_{21} - 0.29x_{22}}}{1 - e^{0.0087 - 0.027x_{11} - 0.5x_{12} - 0.47x_{13} - 1.9x_{14} - 1.96x_{15} - 0.27x_{16} + 0.51x_{21} - 0.29x_{22}}}$$

Where X1 represent the person's education and X2 represents the person's work status.

4.3 Model Interpretation

Variable Name	Variable Category	Odds Ratio	Interpretation
Education of person	**read and write**	0.9727458	Since p-value= 0.918 which is > 0.05 and the C.I includes 1 then, we can say that the odds of not using proper means of garbage disposal among those who read and write is not significantly different from that among illiterate holding the person's work status constant.
Education of person	**basic education**	0.6063856	Since p-value= 0.188 which is > 0.05 and the C.I includes 1 then, we can say that the odds of not using proper means of garbage disposal among those who have a basic education is not significantly different from that among illiterate holding the person's work status constant.
Education of person	**secondary education**	0.6243805	Since p-value= 0.126 which is > 0.05 and the C.I includes 1 then, we can say that the odds of not using proper means of garbage disposal among those who have a secondary education is not significantly different from that among illiterate holding the person's work status constant.
Education of person	**Diploma**	0.1488599	Since p-value= 0.016 which is < 0.05 and the C.I don't include 1 then, we can say that the odds of not using proper means of garbage disposal among those who has diploma education is lower than that of illiterate people by 0.8511 (85.11%) holding the person's work status

			constant.
Education of person	**university degree**	0.1404734	Since p-value= 0.000 which is < 0.05 and the C.I don't include 1 then, we can say that the odds of not using proper means of garbage disposal among those who has university degree education is lower than that of illiterate people by 0.8595 (85.95%) holding the person's work status constant.
Education of person	**higher than university degree**	0.7650854	Since p-value= 0.853 which is > 0.05 and the C.I includes 1 then, we can say that the odds of not using proper means of garbage disposal among those who have a higher than university degree is not significantly different from that among illiterate holding the person's work status constant.
Work status of person	**Self-employed employing others**	1.678876	Since p-value= 0.070 which is > 0.05 and the C.I includes 1 then, we can say that the odds of not using proper means of garbage disposal among those whose work status is self-employed but employing others is not significantly different from that among those who have a waged job holding the person's education constant.
Work status of person	**Self-employed not employing others**	0.7475197	Since p-value= 0.275 which is > 0.05 and the C.I includes 1 then, we can say that the odds of not using proper means of garbage disposal among those whose work status is self-employed but not employing others is not significantly different from that among those who have a waged job holding the person's education constant.
Intercept (1.008743)		The odds of not using proper means of garbage disposal among those who are illiterate and have a waged job is 1.008743	

4.4 The model's significance:

Please note that: if we compared our model to the intercept only model we will found that our model is a significant model since as illustrated in the following output; p-value < 0.05 which is an approval for the model significance.

```
Logistic regression                             Number of obs     =        469
                                                LR chi2(8)        =      48.54
                                                Prob > chi2       =     0.0000
Log likelihood = -291.93167                     Pseudo R2         =     0.0768
```

4.5 Testing the goodness of fit for our model:

$$H_0: our\ model\ has\ a\ good\ fit$$

$$H_1: our\ model\ do\ not\ have\ a\ good\ fit$$

After we conducted our binary logistic model we are going to test the goodness of fit for our model using Person Chi-squared test.

```
. estat gof

Logistic model for Garbage, goodness-of-fit test

       number of observations =       469
 number of covariate patterns =        19
             Pearson chi2(10) =         9.36
                   Prob > chi2 =         0.4986

.
end of do-file
```

The Chi-squared test resulted in a p-value equals to 0.4986 which is greater than alpha=0.05. From this, we can conclude that there is no evidence of poor fit and the model is correctly specified.

4.6 The model's predictive power:

Then we have tested the predictive power of the model using two different techniques:

- Classification Table at two different cutoff points.
- The ROC curve.

4.6.1 The Classification Table:

```
. estat classification

Logistic model for Garbage
```

Classified	True D	True ~D	Total
+	61	45	106
-	128	235	363
Total	189	280	469

```
Classified + if predicted Pr(D) >= .5
True D defined as Garbage != 0
```

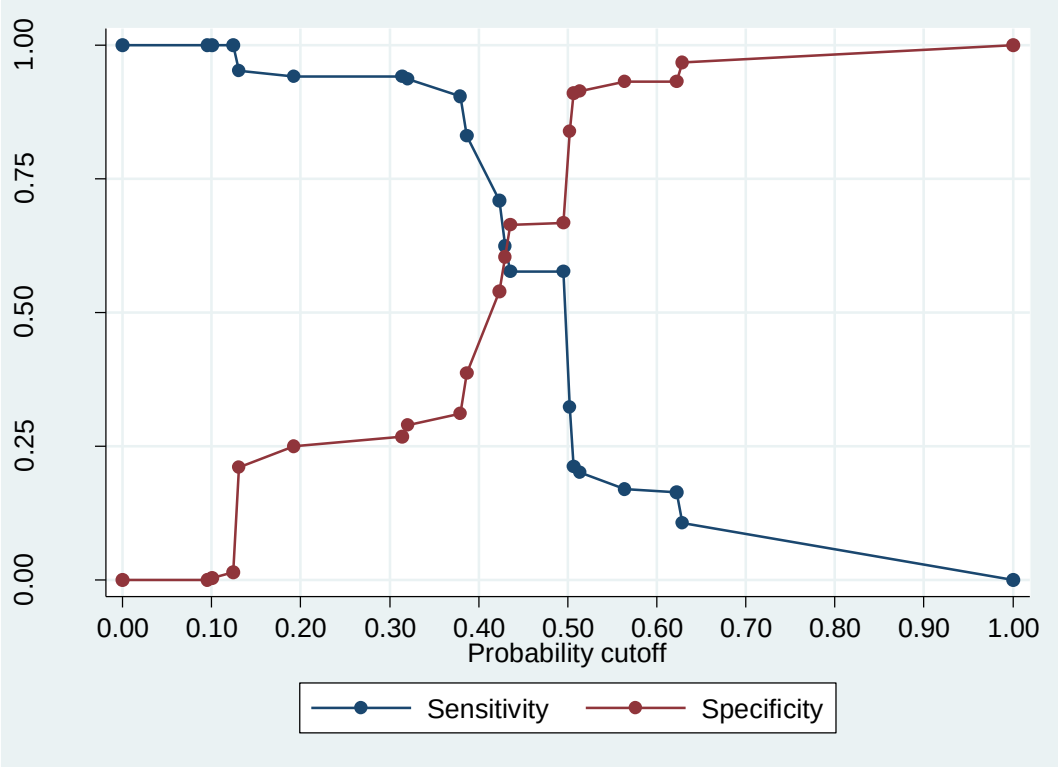
Sensitivity	Pr(+| D)	32.28%
Specificity	Pr(-|~D)	83.93%
Positive predictive value	Pr(D| +)	57.55%
Negative predictive value	Pr(~D| -)	64.74%
False + rate for true ~D	Pr(+|~D)	16.07%
False - rate for true D	Pr(-| D)	67.72%
False + rate for classified +	Pr(~D| +)	42.45%
False - rate for classified -	Pr(D| -)	35.26%
Correctly classified		63.11%

```
.
end of do-file
```

Comment:

Based on the default cutoff point (0.5) we can conclude that the model predictive power is not that good since "Sensitivity" which the probability of correctly specifying a success; the person don't use the proper means of garbage disposal is equal to 32.28 which a very low and unacceptable percentage while "Specificity" which the probability of correctly specifying a failure; the person uses the proper means of garbage disposal is equal to 83.93% which is a very good percentage. We also found that the overall percentage of correct classification is 63.11% which is an acceptable but not that strong. As a conclusion we found that the predictive power of our model at a cutoff point (0.5) is not that good.

Hence we used STATA in order to determine the best cutoff point; we graphed "Sensitivity" and "Specificity" at the different cutoff points and the result was as follows:



We found that we can consider the best cutoff point as approximately 0.42 hence we made another cross classification table at cutoff point=0.42

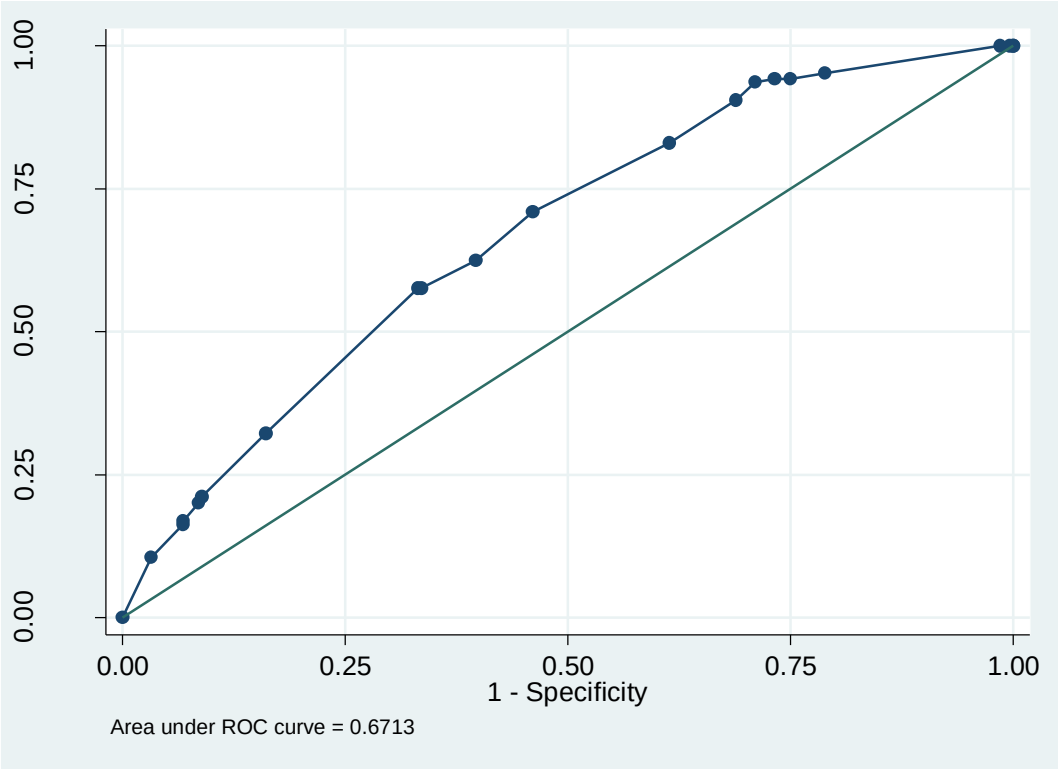
And we found the following:

```
. estat classification, cutoff(0.42)

Logistic model for Garbage

                 -------- True --------
Classified |         D             ~D |       Total
-----------+--------------------------+-----------
     +     |       134            129 |         263
     -     |        55            151 |         206
-----------+--------------------------+-----------
   Total   |       189            280 |         469

Classified + if predicted Pr(D) >= .42
True D defined as Garbage != 0
-------------------------------------------------
Sensitivity                     Pr( +| D)   70.90%
Specificity                     Pr( -|~D)   53.93%
Positive predictive value       Pr( D| +)   50.95%
Negative predictive value       Pr(~D| -)   73.30%
-------------------------------------------------
False + rate for true ~D        Pr( +|~D)   46.07%
False - rate for true D         Pr( -| D)   29.10%
False + rate for classified +   Pr(~D| +)   49.05%
False - rate for classified -   Pr( D| -)   26.70%
-------------------------------------------------
Correctly classified                        60.77%
-------------------------------------------------


.
end of do-file
```

Comment:

Based on the cutoff point (0.42) we can conclude that the model predictive power is not good at the new cutoff point since “Sensitivity” which the probability of correctly specifying a success; the person don’t use the proper means of garbage disposal is equal to 70.90% which is an acceptable and good percentage while “Specificity” which the probability of correctly specifying a failure; the person uses the proper means of garbage disposal is equal to 53.93% which is a low percentage. We also found that the overall percentage of correct classification is 60.77% which is an acceptable but not strong. As a conclusion we found that the predictive power of our model at a cutoff point (0.42) is not good as well.

4.6.2 The ROC curve:



Comment:

It is obvious from the graph that the curve is close to the 45° line which means the model don't have a high overall accuracy. We can also deduce that the model don't have a very good predicative power, as the AUC is equal to 0.6713 (67.13%) which indicates that the model don't have a high true positive rate.

Section (5)

Conclusion/ Recommendation:

To sum up, in this paper we have conducted a study on a sample from Ismailia Governorate citizens regarding how they tend to behave about garbage disposal and whether their education, work status or even activity can affect their behavior. We have concentrated on how proper means of garbage disposal can affect our life. Also, when we conducted our study we found that both education and work status can affect the people's attitude regarding garbage disposal. However, we also see that each country must make an awareness campaign for its citizens to inform them about how important is it to use the proper means of garbage disposal regardless of your education, activity or any other factors you as a citizen must follow the given rules of garbage disposal whoever.

Note: please find the codes on the "Do STATA file" on the following link:

https://drive.google.com/file/d/154JSnQKpyB6dA6-pEHZhV2a3iww9Jr-r/view?usp=sharing

References

Pruss-Ustun, A., Giroult, E., Rushbrook, P., & World Health Organization. (1999). *Safe management of wastes from health-care activities*. World Health Organization.

Waste: a problem or a resource?. European Environment Agency. (2014). Retrieved 25 December 2021, from https://www.eea.europa.eu/publications/signals-2014/articles/waste-a-problem-or-a-resource.

6 Negative Effects of Improper Waste Management. Metropolitantransferstation.com.au. (2017). Retrieved 25 December 2021, from https://www.metropolitantransferstation.com.au/blog/negative-effects-of-improper-waste-management/.

Appendix

```
. doedit "E:\ghofran\stata\h.do"

. use "E:\ghofran\stata\Ismailia.dta"

. mvdecode wrkstat activity, mv(999)
     wrkstat: 59 missing values generated
    activity: 58 missing values generated

. do "C:\Users\DELL\AppData\Local\Temp\STD00000000.tmp"

. tab garbage
```

Proper means of garbage disposal	Freq.	Percent	Cum.
no	203	38.45	38.45
yes	325	61.55	100.00
Total	528	100.00	

```
. tab garbage, nolab

    Proper
   means of
    garbage
   disposal |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        203       38.45       38.45
          1 |        325       61.55      100.00
------------+-----------------------------------
      Total |        528      100.00

. gen Garbage = 0 if garbage == 1
(203 missing values generated)

. replace Garbage=1 if garbage==0
(203 real changes made)

. label define x 0"yes" 1"no"

. label values Garbage x

. tab Garbage, nolab

    Garbage |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        325       61.55       61.55
          1 |        203       38.45      100.00
------------+-----------------------------------
      Total |        528      100.00

.
```

```
.
.
. * 2 way con tab
. tab wrkstat Garbage
```

Work Status of Person	Garbage yes	no	Total
waged job	199	117	316
self employed employi	30	40	70
self employed not em	51	32	83
Total	280	189	469

```
. tab activity Garbage
```

Activity of Person	Garbage yes	no	Total
agriculture	42	46	88
manufacturing	21	10	31
electricity	9	5	14
construction	27	18	45
trade	54	33	87
transportation	35	23	58
finance	8	1	9
services	84	54	138
Total	280	190	470

```
. tab educate Garbage
```

Education of Person	Garbage yes	no	Total
illiterate	53	57	110
read and write	86	79	165
basic education	40	18	58
secondary education	60	36	96
Diploma	15	2	17
university degree	70	10	80
higher than universit	1	1	2
Total	325	203	528

```
.
.
. * test of indep
. * 2 way con tab
. tab wrkstat Garbage, chi2
```

Work Status of Person	Garbage yes	no	Total
waged job	199	117	316
self employed employi	30	40	70
self employed not em	51	32	83
Total	280	189	469

Pearson chi2(2) = 9.7675 Pr = 0.008

```
. tab activity Garbage, chi2
```

Activity of Person	Garbage yes	no	Total
agriculture	42	46	88
manufacturing	21	10	31
electricity	9	5	14
construction	27	18	45
trade	54	33	87
transportation	35	23	58
finance	8	1	9
services	84	54	138
Total	280	190	470

Pearson chi2(7) = 9.6662 Pr = 0.208

```
. tab educate Garbage, chi2
```

Education of Person	Garbage yes	no	Total
illiterate	53	57	110
read and write	86	79	165
basic education	40	18	58
secondary education	60	36	96
Diploma	15	2	17
university degree	70	10	80
higher than universit	1	1	2
Total	325	203	528

Pearson chi2(6) = 43.8818 Pr = 0.000

```
.
. * tset association
. tab wrkstat Garbage, V
```

Work Status of Person	Garbage yes	no	Total
waged job	199	117	316
self employed employi	30	40	70
self employed not em	51	32	83
Total	280	189	469

Cramér's V = 0.1443

```
. tab activity Garbage, V
```

Activity of Person	Garbage yes	no	Total
agriculture	42	46	88
manufacturing	21	10	31
electricity	9	5	14
construction	27	18	45
trade	54	33	87
transportation	35	23	58
finance	8	1	9
services	84	54	138
Total	280	190	470

Cramér's V = 0.1434

```
. tab educate Garbage, gamma
```

Education of Person	Garbage yes	no	Total
illiterate	53	57	110
read and write	86	79	165
basic education	40	18	58
secondary education	60	36	96
Diploma	15	2	17
university degree	70	10	80
higher than universit	1	1	2
Total	325	203	528

gamma = -0.3730 ASE = 0.057

```
. table wrkstat Garbage, by(educate)
```

Education of Person and Work Status of Person	Garbage yes	no
illiterate		
waged job	20	21
self employed employing others	9	20
self employed not employing others	17	9
read and write		
waged job	48	48
self employed employing others	10	11
self employed not employing others	18	16
basic education		
waged job	21	14
self employed employing others	1	2
self employed not employing others	6	1
secondary education		
waged job	43	23
self employed employing others	5	6
self employed not employing others	6	6
Diploma		
waged job	11	2
self employed employing others		
self employed not employing others	3	

university degree		
waged job	55	9
self employed employing others	5	
self employed not employing others	1	
higher than university degree		
waged job	1	
self employed employing others		1
self employed not employing others		

```
. * binary log model
. xi : stepwise, pr(0.05) : logit Garbage i.educate i.activity i.wrkstat
i.educate         _Ieducate_2-8       (naturally coded; _Ieducate_2 omitted)
i.activity        _Iactivity_1-9      (naturally coded; _Iactivity_1 omitted)
i.wrkstat         _Iwrkstat_1-3       (naturally coded; _Iwrkstat_1 omitted)
                      begin with full model
p = 0.8993 >= 0.0500  removing _Ieducate_3
p = 0.8896 >= 0.0500  removing _Iactivity_4
p = 0.9000 >= 0.0500  removing _Iwrkstat_3
p = 0.6674 >= 0.0500  removing _Ieducate_8
p = 0.5839 >= 0.0500  removing _Iactivity_9
p = 0.3379 >= 0.0500  removing _Iactivity_8
p = 0.3083 >= 0.0500  removing _Iactivity_7
p = 0.2823 >= 0.0500  removing _Iactivity_5
p = 0.1514 >= 0.0500  removing _Ieducate_4
p = 0.1866 >= 0.0500  removing _Ieducate_5
p = 0.1248 >= 0.0500  removing _Iactivity_3
p = 0.0951 >= 0.0500  removing _Iactivity_6

Logistic regression                             Number of obs     =        469
                                                LR chi2(3)        =      43.56
                                                Prob > chi2       =     0.0000
Log likelihood = -294.42248                     Pseudo R2         =     0.0689
```

```
Logistic regression                             Number of obs     =        469
                                                LR chi2(3)        =      43.56
                                                Prob > chi2       =     0.0000
Log likelihood = -294.42248                     Pseudo R2         =     0.0689
```

Garbage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Iwrkstat_2	.654064	.2695936	2.43	0.015	.1256703	1.182458
_Ieducate_7	-1.719438	.3728508	-4.61	0.000	-2.450212	-.9886639
_Ieducate_6	-1.693715	.76426	-2.22	0.027	-3.191637	-.1957934
_cons	-.2521947	.1125377	-2.24	0.025	-.4727645	-.0316249

```
. logit Garbage i.educate i.wrkstat

Iteration 0:   log likelihood = -316.20142
Iteration 1:   log likelihood = -292.40079
Iteration 2:   log likelihood = -291.93369
Iteration 3:   log likelihood = -291.93167
Iteration 4:   log likelihood = -291.93167

Logistic regression                             Number of obs     =        469
                                                LR chi2(8)        =      48.54
                                                Prob > chi2       =     0.0000
Log likelihood = -291.93167                     Pseudo R2         =     0.0768
```

Garbage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
educate						
read and write	-.0276325	.2684826	-0.10	0.918	-.5538487	.4985837
basic education	-.5002392	.3796114	-1.32	0.188	-1.244264	.2437854
secondary education	-.4709953	.3079574	-1.53	0.126	-1.074581	.13259
Diploma	-1.90475	.7899053	-2.41	0.016	-3.452936	-.356564
university degree	-1.962737	.4245876	-4.62	0.000	-2.794914	-1.130561
higher than university degree	-.2677678	1.44335	-0.19	0.853	-3.096682	2.561146
wrkstat						
self employed employing others	.5181247	.2859932	1.81	0.070	-.0424116	1.078661
self employed not employing others	-.2909946	.2664209	-1.09	0.275	-.81317	.2311808
_cons	.0087054	.2404112	0.04	0.971	-.4624919	.4799027

```
. logit Garbage i.educate i.wrkstat, or

Iteration 0:   log likelihood = -316.20142
Iteration 1:   log likelihood = -292.40079
Iteration 2:   log likelihood = -291.93369
Iteration 3:   log likelihood = -291.93167
Iteration 4:   log likelihood = -291.93167

Logistic regression                             Number of obs     =        469
                                                LR chi2(8)        =      48.54
                                                Prob > chi2       =     0.0000
Log likelihood = -291.93167                     Pseudo R2         =     0.0768

------------------------------------------------------------------------------
             Garbage | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
             educate |
          read and write |   .9727458   .2611653    -0.10   0.918     .5747336    1.646388
        basic education |   .6063856   .2301909    -1.32   0.188      .288153     1.27607
     secondary education |   .6243805   .1922826    -1.53   0.126     .3414409    1.141782
             Diploma |   .1488599   .1175852    -2.41   0.016     .0316526    .7000777
          university degree |   .1404734   .0596433    -4.62   0.000     .0611202    .3228521
higher than university degree |   .7650854   1.104286    -0.19   0.853     .0451989    12.95065
                     |
             wrkstat |
self employed employing others |   1.678876   .4801471     1.81   0.070     .9584752    2.940739
self employed  not employing others |   .7475197   .1991549    -1.09   0.275     .4434501    1.260087
                     |
               _cons |   1.008743   .2425132     0.04   0.971     .6297125    1.615917
------------------------------------------------------------------------------

.
end of do-file
```

```
Logistic model for Garbage, goodness-of-fit test

       number of observations =       469
 number of covariate patterns =        19
             Pearson chi2(10) =      9.36
                  Prob > chi2 =    0.4986

. estat classification

Logistic model for Garbage

                 ---------- True ----------
Classified |         D              ~D      |      Total
-----------+--------------------------------+-----------
     +     |        61              45      |        106
     -     |       128             235      |        363
-----------+--------------------------------+-----------
   Total   |       189             280      |        469

Classified + if predicted Pr(D) >= .5
True D defined as Garbage != 0
--------------------------------------------------------------
Sensitivity                     Pr( +| D)   32.28%
Specificity                     Pr( -|~D)   83.93%
Positive predictive value       Pr( D| +)   57.55%
Negative predictive value       Pr(~D| -)   64.74%
--------------------------------------------------------------
False + rate for true ~D        Pr( +|~D)   16.07%
False - rate for true D         Pr( -| D)   67.72%
False + rate for classified +   Pr(~D| +)   42.45%
False - rate for classified -   Pr( D| -)   35.26%
--------------------------------------------------------------
Correctly classified                        63.11%
--------------------------------------------------------------
```

```
. lsens

.
.
end of do-file

. do "C:\Users\DELL\AppData\Local\Temp\STD00000000.tmp"

. estat classification, cutoff(0.42)

Logistic model for Garbage

              -------- True --------
Classified |         D             ~D |       Total
-----------+--------------------------+-----------
     +     |       134            129 |         263
     -     |        55            151 |         206
-----------+--------------------------+-----------
   Total   |       189            280 |         469

Classified + if predicted Pr(D) >= .42
True D defined as Garbage != 0
```

```
-------------------------------------------------
Sensitivity                     Pr( +| D)   70.90%
Specificity                     Pr( -|~D)   53.93%
Positive predictive value       Pr( D| +)   50.95%
Negative predictive value       Pr(~D| -)   73.30%
-------------------------------------------------
False + rate for true ~D        Pr( +|~D)   46.07%
False - rate for true D         Pr( -| D)   29.10%
False + rate for classified +   Pr(~D| +)   49.05%
False - rate for classified -   Pr( D| -)   26.70%
-------------------------------------------------
Correctly classified                        60.77%
-------------------------------------------------

. lroc

Logistic model for Garbage

number of observations =      469
area under ROC curve   =   0.6713

.
end of do-file
```

```
. tab educate
```

Education of Person	Freq.	Percent	Cum.
illiterate	110	20.83	20.83
read and write	165	31.25	52.08
basic education	58	10.98	63.07
secondary education	96	18.18	81.25
Diploma	17	3.22	84.47
university degree	80	15.15	99.62
higher than university degree	2	0.38	100.00
Total	528	100.00	

```
. tab wrkstat
```

Work Status of Person	Freq.	Percent	Cum.
waged job	316	67.38	67.38
self employed employing others	70	14.93	82.30
self employed not employing others	83	17.70	100.00
Total	469	100.00	