Demographic Research On the Socio Economic Background of Students of the Ecological University of Bucharest

Ph.D. Janina Mihaela MIHĂILĂ (janinamihaelamihaila@yahoo.it)
Ecological University of Bucharest

ABSTRACT

The paper describes a socio demographic and economic research performed on the first year students at the Ecological University of Bucharest, where we are focusing on understanding and investigating the conditions inside the families and the social environment in the home towns of these students. This research is a key in understanding the correlations between the socio-economic conditions inside the family geographical area and the actual career options and decisions of the newly admitted students to our faculties.

Key Words: *correlation, R programming language, demographic research, universities.*

JEL Classification: *I10, I20, I23, I25, J10.*

INTRODUCTION

The importance of demographic studies in modern societies is just huge. All sort of entities, from multinational companies, to retailing giants, to government and city municipalities resort to them, in order to be able to correctly assess and estimate the impact of a certain factor on a specific segment of the population.

For correctly taking all decisions about managing the advertorial system of the University, faculties and studies program, we must evaluate the socio economic premises of the geographical and demographical area of our potential students.

The present paper focuses on 3 important aspects:

- economic
- social - demographical areas
- cultural

For this research, our aim was to investigate the dependencies between the socio demographic factors such as:

- the family's income
- regional development and specific opportunities
- family's level of education.

There is a clear connection between the socio-demographic environment from where a student comes, and his or her later career development and future opportunities.

At the beginning of the 20th century Sigmund Freud remarked the profound connection between the development in the early years of childhood and a wide variety of options in adult life.

We start by choosing a representative sample: in our case we perform the study on the entire lot of students in the first year that joined the Ecological University in Bucharest.

They may be considered as a representative sample, as we have people from all the regions of Romania and from a wide variety of social backgrounds.

Each student was given a specific questionnaire, without any prior notice and asked to fill it in one hour. It contained questions that helped us understand and accurately represent their origin from a social, cultural and regional point of view.

In this way, we have understood the complexity of factors that had led them to join this specific university instead of others.

All answers were anonymous.

The first step in our project was to have the distribution characterized inside the sample considered.

This study has been conducted on a sample of 249 students, having the following structure:

Gender and age defined structure of the sample.

Table 1

Gender / Age (years)	F	M	Σ	%
18 - 29	54	37	91	36.55
30 - 40	76	43	119	47.79
> 40	19	20	39	15.66
Σ	**149**	**100**	**249**	**100**

We had a very high response rate for these questionnaires of 90,55 %, as we distributed to students a total of 275 surveys and we received a total of 249 of them back ,completed with all information required; this very high percentage was also on a large extent due to the fact that the students were provided with special designated collector box for the quesionnaires easy return.

Upon receiving, the results were processed with dedicated software and we were able to have a more detailed image of the correlations of the students's choice for a specific faculty.

Sample's structure compared to the district's population number

Table 2

District	Population	Weight	Respondents	%
D1	1628426	0.25	67	**26.90**
D2	301425	0.05	23	**9.24**
D3	316652	0.05	10	**4.02**
D4	674903	0.1	6	**2.40**
D5	680945	0.11	24	**9.64**
D6	353481	0.05	5	**2.01**
D7	654870	0.1	11	**4.42**
D8	540508	0.08	13	**5.22**
D9	211622	0.03	18	**7.23**
D10	374240	0.06	37	**14.86**
D11	393340	0.06	16	**6.43**
D12	339510	0.05	19	**7.63**
Σ	**6469922**	**1**	**249**	**100**

INPUT DATA

Respondent's structure by educational level and district

Table 3

Family's level of education / District of Birth	Middle school	High school graduation	Bachelor's/ master's degree	Σ
D1	0	23	44	67
D2	2	9	12	23
D3	0	3	7	10
D4	0	1	5	6
D5	1	7	16	24
D6	0	2	3	5
D7	1	5	5	11
D8	0	4	9	13
D9	3	8	7	18
D10	3	15	19	37
D11	1	10	5	16
D12	2	12	5	19
Σ	**13**	**99**	**137**	**249**
%	**5.22**	**39.76**	**55.02**	**100**

Table 4. Respondent's family monthly income by district of birth.

Family monthly income (RON) / District of Birth	Under 2000	2001 – 4000	4001 – 6000	6001 –10,000	Above 10,000
D1	32	14	19	1	1
D2	18	2	3	0	0
D3	7	2	1	0	0
D4	5	1	0	0	0
D5	17	3	3	1	0
D6	3	1	1	0	0
D7	9	1	0	0	1
D8	11	2	0		0
D9	13	5	0	0	0
D10	16	15	5	1	0
D11	10	4	2	0	0
D12	6	11	2	0	0
Σ	**147**	**61**	**36**	**3**	**2**
%	**59.04**	**24.50**	**14.46**	**1.20**	**0.80**

Remark. NUTS 3 statistical regions of the European Union.

There are three levels defined in the Nomenclature of Territorial Units for Statistics (NUTS). The above category refers to regions belonging to the third level (NUTS 3, also known as NUTS III), which is largely used by Eurostat and other European Union bodies.

Table 5. Respondent's structure by demographic areas and district of birth.

Demographical area of birth / District of Birth	Rural	Intermediar (suburban)	Urban	Σ
D1	5	28	34	**67**
D2	7	1	15	**23**
D3	2	6	2	**10**
D4	1	3	2	**6**
D5	4	8	12	**24**
D6	0	0	5	**5**
D7	3	2	6	**11**
D8	4	5	4	**13**
D9	2	7	9	**18**
D10	9	9	19	**37**
D11	6	7	3	**16**
D12	8	4	7	**19**
Σ	**51**	**80**	**118**	**249**
%	**20.48**	**32.13**	**47.39**	**100.00**

STATISTICAL ANALYSIS

Calculation of the correlation coefficient (Pearson).

We compute the correlation coefficient or Pearson coefficient interpreted according to Colton empirical rules (Theodore Colton, Professor Boston University).

The correlation coefficient or Pearson coefficient is an indicator independent of the units of measure of the two variables

$$r = \frac{COV(X,Y)}{Sx \cdot Sy}$$

where S_X and S_Y represent the standard deviations for the X and Y series respectively.

Indicators definition.

In order to evaluate in an exact numerical expression the possibility and the individual decision to enroll in a higher education institution, we define the following indicators:

1. accessibility to higher education services indicator ACIHE as an expression of the level of accessibility to higher education programs.
2. aspiration to higher education services indicator ASIHE expressing the extent in which the individual is determined to enroll in a higher education study program.

Calculation Formula

For answers Q7- Q14 items are assigned with numerical values corresponding to different upward variants, which were marked with values from 0 to n.

The following formulas are being used:

1. ACIHE indicator is calculated as the sum of the corresponding values given by respondents to the items: Q10, Q12, and Q14 in the questionnaire.
2. ASIHE indicator is calculated as that sum of the corresponding values recorded by respondents to the items: Q7, Q8, Q9 from questionnaire.

TECHNIQUES AND STATISTICAL METHODS OF CALCULATION USED.

The R programming language was used for the input, storage and calculation of data in order to better familiarize the students with this programming language and develop their skills during this process.

In developing this application, the following instructions or package instructions (scripts) were used:

a. for data table creation we used:

- the *read.table* command, but also the alternative *read. table("clipboard")* command may be used if the data are saved in .xsl format

```
tabel s (student)
s <- read.table(text=
“District      Middle     High    Bachelor Master
D1                7          54          25          5
D2      3          16           18           2
D3      1          10           12           4
D4                0          10          19          1
D5                1          1            0           0
“, header=TRUE)
s
```

matrix type entities through: *cbind()* or *rbind()* functions
data.frame objects

b. in performing the various numerical calculations on data we used:
- storing it in objects: variables, vectors, matrixes, tables or data frames
- the interactive way

c. calculating the Pearson correlation coefficient for the corresponding data sets with the statistical correlation function

d. viewing the processing objects that have been used and the search results:
- a pie chart showing the distribution of the income vector

```
pie(stud, main=”Student’s income in District 1 (RON)”, col=rainbow(length(stud)), labels=c(“<2000”,”2000-4000”,”4000-6000”,”6000-8000”,”>8000”))
legend(1.25,0.5,c(“<2000”,”2000-4000”,”4000-6000”,”6000-8000”,”>8000”), cex=0.8, fill=colors)
```

- histogram showing the distribution of the students vector

```
brk <- c(0,3,4,5,6,10,16)
hist(Students, col=heat.colors(length(brk)), breaks=brk, xlim=c(0,max_num), right=F, main=” Student Density”, las=1, cex.axis=0.8, freq=F)
```

- a line chart with two axes and a legend in which we compute the y-axis values using the max function so any changes of these data will be automatically reflected in the graph.

```
High school<- c(23, 9, 3, 1, 7)
Bachelor <- c(44, 12, 7, 5, 16)
g_range <- range(0, High school,Bachelor)
plot(Highschool, type=”o”, col=”blue”, ylim=g_range, axes=FALSE, ann=FALSE)
axis(1, at=1:5, lab=c(“D1”,”D2”,”D3”,”D4”,”D5”))
axis(2, las=1, at=10*0:g_range[2])
box()
lines(Bachelor, type=”o”, pch=22, lty=2, col=”red”)
title(main=”CHART3”, col.main=”red”, font.main=4)
title(xlab=”Districts”, col.lab=rgb(0,0.5,0))
title(ylab=”Number”, col.lab=rgb(0,0.5,0))
legend(1, g_range[2], c(“Highschool”,”Bachelor”), cex=0.8, col=c(“blue”,”red”), pch=21:22, lty=1:2);
```

RESULTS OBTAINED AND THEIR INTERPRETATION

Indicator calculation

Indicator calculation is performed using the R system, thus speeding up the process, using the following code:

```
i <- read.table(text=
“Respondent    VQ6  VQ7    VQ8     VQ9
R1                  7     2       8       5
R2                  3     2       6       2
R3                  1     1       3       4
R4                  0     0       2       1
R5                  1     1       5       0
          “, header=TRUE)
i
i$VQ7
i$VQ8
ACIHE=i$VQ7+i$VQ8
ACIHE
```

The output is showed on the running window.

Indicator calculation

Fig. 1

```
  Respondent VQ6 VQ7 VQ8 VQ9
1         R1   7   2   8   5
2         R2   3   2   6   2
3         R3   1   1   3   4
4         R4   0   0   2   1
5         R5   1   1   5   0
>
> i$VQ7
[1] 2 2 1 0 1
> i$VQ8
[1] 8 6 3 2 5
>
> ACIHE=i$VQ7+i$VQ8
> ACIHE
[1] 10  8  4  2  6
>
```

Calculating correlations

We compute the Pearson correlation coefficient of the values recorded from the respondents to the items Q7 and Q8.

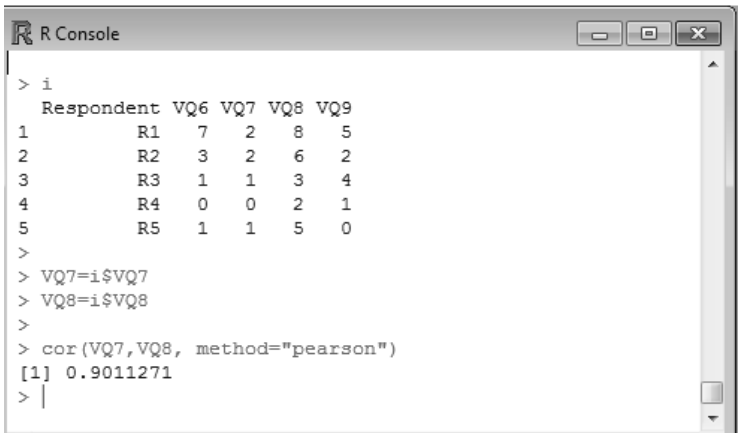
For this we define a table and use the correlation between two columns of the table: VQ7 and VQ8 considered as vectors in which we stored the values corresponding to the responses.

```
VQ7=i$VQ7
VQ8=i$VQ8
cor(VQ7,VQ8, method="pearson")
```

Following the application of the calculation formulas and statistical functions, the following results were displayed in the R window:

Calculating correlation

Fig. 2

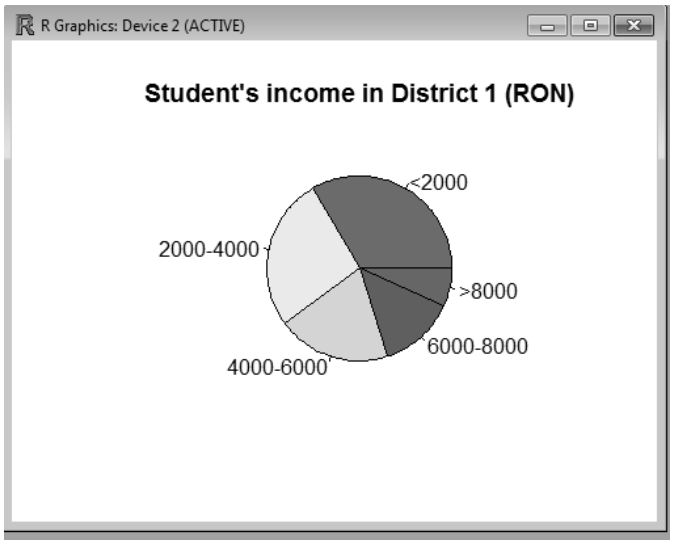


Pearson correlation coefficient is 0.91 which indicates a very good association or correlation between the level of education of the respondent and his/her family (a correlation coefficient greater than 0.75 or less than - 0.75).

Displaying the results.

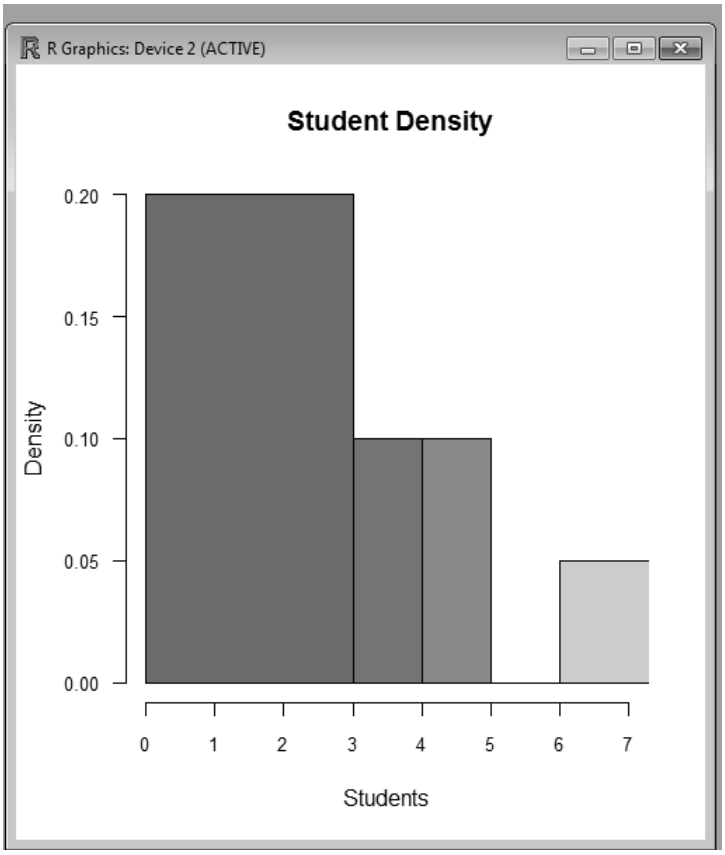
Student's income in District 1 (RON)

Fig. 3



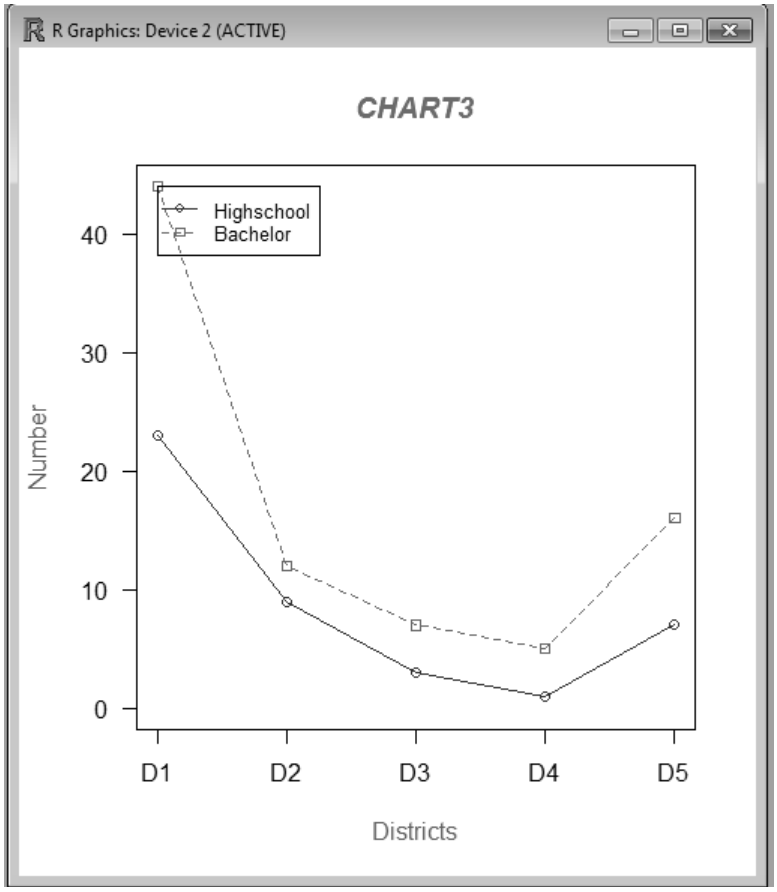
Student Density

Fig. 4



Number of students by educational profile (High School and Bachelor) in districts

Fig 5



CONCLUSIONS

The aim of this paper has been to provide a document to be read and understood by students, their families and also to be accessible for a larger audience. We must keep into account that the target audience does not have deep statistical and IT knowledge, but are nevertheless very keen in knowing the results.

We have wanted to share this knowledge with the wider public and to have it used and further extended in future studies for a didactical purpose,

mostly because it has practical uses in fields such as economics, technics, social sciences and more.

We are strongly encouraging the students to continue working on this area and further develop these knowledge.

The scope of developing such an application is on various areas, such as:

- didactical, to provide practical examples for these theoretical notions and concepts;
- organizational, so that both teachers and students may better organize their teaching/learning activities together;
- practical, for developing the skills and abilities that the students will need for successfully building up their future career paths;
- management, in promoting an authentic university quality management system;
- personal development, through self-knowledge and self-awareness of their own individual motivations and educational aspirations.

RECOMMENDATIONS AND DIRECTIONS FOR FURTHER DEVELOPMENT

We regard as useful and welcomed the development of case study themes, either individual or in group, (which will form small working groups), leading to the built up of scientific and research skills, helpful for elaborating the undergraduate thesis and diploma dissertation preparation.

References

1. Colton T., Freedman L. S. and Johnson A. L.. Statistics in medicine, Vol. 1, No. 1, January, 1982, New York, John Wiley and Sons.
2. Colton, T. Statistics in Medicine. Little Brown and Company, New York, NY 1974.
3. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
4. Venables W. N., Smith D. M., R Core Team. An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. 2012.
5. Murrell, P. (2005) R Graphics. Chapman & Hall/CRC Press.
6. http://cran.r-project.org/doc/manuals/R-intro.html
7. http://www.r-project.org/

ANNEX

Sample questionnaire

Q1. Please indicate your sex F M
Q2. Please indicate your year of birth ……
Q3. Do you have siblings Yes/No and if yes, how many……
Q4. Ethnic origin ……..
Q5. Marital status …….
Q6. Family background……..
Q7. Your level of education (highest level of education)

A. Middle school
B. High school graduation
C. Bachelor's degree or master's degree

Q8. What is the highest level of education your parents have completed? *If currently enrolled, highest degree received.*

A. No schooling completed
B. Nursery school to 8th grade
C. Some high school, no diploma
D. High school graduate, diploma or the equivalent (for example: GED)
E. Some college credits, no degree
F. Trade/technical/vocational training
G. Bachelor's degree
H. Master's degree
I. PhD degree

Q9. Grandparents highest level of education ……
Q10. Are you employed? A. full time B. part time C. unemployed
Q11. Employment contract: A. none B. temporary C. permanent
Q12. The time schedule allows you to be present in class during teaching/ learning activities?

A. weekends only B. partially C. completely

Q13. What kind of area were you raised in?
A. rural B. small town C. suburban D. urban
Q14. Please report an estimate of your family's total monthly income:

A. Under RON 2.000
B. Between RON 2,001 and 4.000
C. Between RON 4,001 and 6.000
D. Between RON 6,001 and 10.000
E. More than RON 10.000

Q15. What were your hometown extra-curricular activities available during your

years in school? Please list the ones you followed – if any.....
Q16. Did your teaching institution in your hometown provided access to the following activities:

A. library

B. gym

C. IT lab with internet

D. uptodate knowledge on student grants, exchange programs and other facilities

Q17. Did your home town had:

A. Museums B. Cinemas C. Theatres D. Opera houses E. Stadiums