Statistical Literacy and Attitudes Towards Statistics of Romanian Undergraduate Students

Cimpoeru, Smaranda and Roman, Monica

The Bucharest University of Economic Studies, Department of Statistics and Econometrics, The Bucharest University of Economic Studies, Department of Statistics and Econometrics

1 March 2018

Statistical Literacy and Attitudes Towards Statistics of Romanian Undergraduate Students[1]

Smaranda CIMPOERU[a] , Monica ROMAN[b]

Abstract

Statistical literacy is the capacity to challenge statistics encountered in everyday life (Gal, 2002) and has become a key competence for the entire work-force in today's data-driven society. However, teaching statistics to non-economists has some particularities determined by their anxiety towards the subject. In order to have a better approach in teaching Statistics to this type of audience, the paper provides an analysis regarding students' initial level of statistical literacy, the attitudes and beliefs towards Statistics, in the case of undergraduate students of the "Applied Modern Language" program under Bucharest University of Economic Studies. Using the models developed by Gal (2002) and Watson (2003), the results prove that basic statistical literacy skills like graph analysis or table reading are sound, while mathematical level is well below average, correlated with a high anxiety regrading mathematics and preconceptions about complexity of statistics.

Keywords: Statistical literacy; Statistics education; undergraduate non-economists students; attitude towards statistics; teaching and learning statistics.

JEL Classification: A22, C19, I23.

Authors' Affiliation

[a] – PhD., Lecturer, Bucharest University of Economic Studies, Faculty of Cybernetics, Statistics and Informatics, Department of Statistics and Econometrics; smaranda.cimpoeru@csie.ase.ro (corresponding author)

[b] – PhD, Professor, Bucharest University of Economic Studies, Faculty of Cybernetics, Statistics and Informatics, Department of Statistics and Econometrics

[1] A preliminary version of this paper was presented at The 11th International Conference of Applied Statistics, Brasov (Romania), June 2017. The authors thank participants for the valuable feedback.

1. Introduction

Initially considered as the science (branch of mathematics by most formal definitions) dealing with the collection, classification, analysis, and interpretation of numerical facts or data[2], statistics has evolved over time to be nowadays "the science of learning from data" (Horton, 2015). This beautiful definition of statistics has been promoted by the American Statistical Association[3], but also by other important statistical associations[4].

It is widely known that statistics can be associated with various sciences like biology, chemistry, meteorology, physics or sociology. To add more to this point, one can argue that statistics is ubiquitous since it can be found in every aspect of our daily lives, like health, economics and finance, weather, business, etc.

Consequently, mastering statistical thinking and statistical concepts grants powerful instruments to improve our lives and make the best decisions. One of the essential characteristics of the world we are living in is that data is omnipresent and easily accessible, thus those who know how to use, analyse and make sense of data will have the competitive advantage. The worldwide spread of statistical data and tools is part of the Data Revolution (Ridgway, 2016) and it has a deep impact on what statistics means nowadays, but also on what and how is to be taught of statistics.

Yet, despite the importance of statistical literacy and statistical reasoning in various life contexts, the subject is underrated in curriculum frameworks both in secondary / high school, as well as in undergraduate courses. Usually, the focus in courses associated with Introductory Statistics is mainly on the mathematical context underpinning the statistical concepts and regrettably low emphasize is put on the practical aspects of these concepts.

A particular challenge in teaching statistics is dealing with students who do not have a sound mathematical background. This is the case for the Introductory Statistics Course[5] of the second year students in the Applied Modern Language program from the Bucharest University of Economic Studies.

The Applied Modern Language undergraduate program is organized by the Faculty of International Business and Economics of the Bucharest University of Economic Studies. It is

[2] https://www.merriam-webster.com/dictionary/statistics
[3] http://www.amstat.org/
[4] http://www.wldofstatistics.org/
[5] The official name of the discipline is "Research Methods in Social Sciences: Quantitative Research", but the course's content is similar to that used in an Introductory Statistics course.

a particular undergraduate program since it introduces the study of modern languages in the global economic context. Besides communication and specific language/translation skills, the future graduates should also master business and entrepreneurship skills, thus including basic statistical analysis. Since basically the program is intended essentially for students with a philological profile, most of them do not have a solid mathematical background.

Moreover the authors' experience with teaching Introductory Statistics to second year students revealed a form of anxiety associated with the statistics discipline and some misconceptions at the beginning of the semester. All of these triggered the necessity of a specific approach in teaching the introductive notions of statistics, so that the students could acquire the necessary competences of a statistical literate graduate.

Prior to developing a specific statistics pedagogy for the non-economists students described above, a deeper understanding of the audience is demanded. We try to answer the following questions:

- What is the students' current level of statistical literacy? (before taking the Introductory Statistics course) What are their strengths and weaknesses?
- What could be the potential sources of variation for students' level of statistical literacy? Are gender or the high-school profile determinants for the level of statistical literacy?
- What are the students' beliefs and attitudes towards statistics? Is there an enhanced anxiety towards the subject?
- Is there a relationship between the attitude towards statistics and the level of statistical literacy (as determined from the study or even self-assessed?)

In order to answer the research questions, a questionnaire is developed following Gal's model (2002) and (to a lower extent) Watson model (2003) and it is tested in the case of the Applied Modern Language program undergraduate students. The questionnaires analysis reveals the strengths and weaknesses of the students' current statistical literacy level and provides significant insight regarding their attitude and beliefs towards statistics. The considerable information provided by this analysis could be further used to shape teaching and learning activities to be used in the Introductory Statistics course for non-economist students.

The paper contributes to the existing literature on teaching Statistics in several ways: it adapts two models provided by Gal (2002) and Watson (2003) in order to provide a useful tool for Statistics professors interested in assessing both students' statistical literacy and their

attitudes towards Statistics. More than that, the proposed model is applied to Romanian students, being to our knowledge the first attempt to assess the statistical literacy in Romanian context. Going beyond descriptive analysis, the paper also evaluates the possible correlation between the level of statistical literacy and attitudes toward Statistics.

The outline of the paper will be presented briefly. Section 2 details the importance and relevance of statistics for all type of professionals, including non-economists. Section 3 introduces the methodology used, the questionnaire design and the sample description. The Results and Discussion section (Section 4) is structured in four sub-sections, each addressing one of the research questions as mentioned above. The last part of the paper concludes the main findings.

2. The role of Statistical literacy in the professional life

Statistical literacy is broadly defined as the ability to *"interpret, critically evaluate, and communicate about statistical information and messages"* (Gal, 2002). According to this definition, the term "statistical literacy" covers two interrelated components. The first component deals with people's ability to interpret and critically evaluate statistical information, stochastic phenomena or data-based arguments which can be encountered in different contexts, while the second one with their ability to discuss or communicate their opinions regarding the statistical information or to formulate critiques for the conclusions they do not support.

Briefly, statistical literacy is the capacity to challenge statistics encountered in everyday life. It enables us to "*consume and critically digest the wealth of information being produced in today's society*" (Rumsey, 2002). Thus, statistical literacy helps one ask better questions and improve judgment, as well as decision making.

Arguably, statistical literacy is a key ability in a society where data, variation and chance are ubiquitous. Since most adults are consumers and not producers of statistical information, the skills associated with statistical literacy could be activated in various contexts. For instance, it helps people to be aware of trends and social phenomena evolution: crime rates, spread of diseases, environment changes, employment trends, political or social polls results. What is more, it supports people improving their decisions when confronted with chance-based situations such as buying lottery tickets, understanding risks associated with certain

diseases or diagnostics, evaluating financial risks. Consequently, it makes people informed consumers of data and that is way it becomes vital to have statistically literate citizens.

"There has never been a more exciting time to be involved in statistics" argues Ridgway (2016). Indeed, the so called data revolution (part of the larger Fourth Industrial Revolution) has transformed the knowledge landscape and has brought new demands for the labour market. A report put together by McKinsey[6] in 2011 regarding the changes brought by Big data to the economy highlights the shortage of talent, "particularly of people with deep expertise in statistics and machine learning". They also draw attention to the difficulty in training this type of employees, and the rather long timeframe necessary to grow the skills of dealing with big data. It is argued in this report that managing big data is mainly about asking the right questions and using effectively the insights gained by the big data analysis. In fact, the skills of managing big data demand nothing else but understanding the main statistical ideas and attaining a highly developed statistical thinking.

What is more, a study on the future of jobs from World Economic Forum[7] emphasizes that "an ability to work with data and make data-based decisions will become an increasingly vital skill across many job families" (p. 21). Employers seek workforce capable of doing data analysis and mastering visualization techniques of exponentially growing amount of data. Thus, these skills are not affiliated to a single profession (for instance, a statistician), but are and will be required for a variety of jobs, since data revolution is present in all the fields of the society.

It follows that the graduates from an Applied Modern Language program, having the characteristics highlighted previously, would also have strong benefits from acquiring the statistical literacy competencies. Consequently, considering employers' current requests, mastering statistical analysis could ease the process of their integration on the labour market.

3. Methodology and data description

3.1 The proposed model and questionnaire design

The questionnaire was designed to answer the research questions detailed in the paper's introductory section. The questionnaire was distributed at the beginning of the "Introductory

[6] http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

[7] https://www.weforum.org/reports/the-future-of-jobs

Statistics" course (March 2017) to the undergraduate students in the "Applied Modern Languages in Economics" program, in their second year of study (second semester). This is the first Statistics course students are taking in their undergraduate program, but considering the knowledge acquired in elementary school and in high-school, it is expected that students should have the necessary expertise to solve the items proposed in the questionnaire.

The questionnaire was designed following Gal's model (Gal, 2002) of the knowledge and processes "that should be available to adults and to learners graduating from schools or colleges so that they can comprehend, interpret, critically evaluate and react to statistical messages encountered in reading contexts" (Gal, 2002, pg. 3). The model was applied for similar purposes in other studies. For instance, Yotongyos et al. (2014) use the model to assess the level of statistical literacy among undergraduate students in Thailand, while Budgett and Pkamkuch (2010) adapted the model to evaluate an undergraduate course on statistical skills (New Zeeland).

Apart from Gal's model, another extensive study (Watson, 2003) deals with developing a hierarchical construct associated with statistical literacy. For the six levels of understanding identified (Idiosyncratic, Informal, Inconsistent, Consistent non-critical, Critical, Critical mathematical), the item-questions used are given as examples together with the appropriate level of understanding and skills assessed. Since the six levels of understanding can be mapped to the components described in Gal's model, some of the items proposed by Watson were also used in the present questionnaire, as it will be outlined below.

However, since the design of the questionnaire used in this study followed the methodology proposed by Gal (2002), in what follows the main elements of the original model will be considered briefly.

The model has seven elements, grouped on two main components:

- The knowledge component comprised of five cognitive elements:
    - Literacy skills
    - Statistical knowledge
    - Mathematical knowledge
    - Context knowledge
    - Critical question
- Dispositional elements
    - Critical Stance
    - Beliefs and attitudes

The items included in the questionnaire are constructed on this framework. In table 1 below, the items proposed for assessing each level are shortly described.

Table 1 – Competences and items proposed in the questionnaire

Level in Gal's model	**Competences and skills to be attained**	**Proposed items**
Knowledge component		
Literacy skills (Level 1 – L1)	Being able to identify, interpret, read and use information given in lists, tables, charts, graphical displays	Item L1.1: graphical display (histogram) with 3 questions assessing the simple read of information from the graph Item L1.2: contingency table followed by 3 questions to verify the correct use / identification of the information in the table
Statistical knowledge (Level 2 – L2)	S1. Knowing why data are needed and how data are produced; ideas about sampling; representativeness S2. Recognize uncommon tabular or graphical displays, "reading beyond data"; S3. Understanding basic notions of probability, randomness; S4. Have some sense of how data are analysed and conclusions reached.	Item L2.1: decision of buying a new car based on information from different sources – identify most accurate source of data, draw conclusions (S1, S4); Item L2.2: define and give example of "sample" (S1, S3); Item L2.3: give examples of things that happen in a "random" way (S3); Item L2.4: identify the most representative sample from several examples; sources of bias (S3, S4); Item L2.5: estimate the number of fish in a lake – use simple probability calculus (S3); Item L2.6: identify correctly the information from a population age pyramid (S2);
Mathematical knowledge (Level 3 – L3)	Being able to use simple mathematical procedures and computations used to determine percentages or averages; identify the difference between the mean and the median, the advantages of the latter; calculate and correctly interpret percentages, ratios;	L3.1: calculate mean and median for a small set of numbers; decide which one is better to use and why L3.2: transform percentages in absolute values (imbricate percentages); calculate percentages out of the total (ratios not to be summed up); L3.3: use the simple arithmetic average formula to determine the sum of the observed values; L3.4: calculate the price of a product before a discount was applied.
Dispositional Elements		
Critical stance	Questioning attitude towards quantitative messages that could be misleading or biased.	A set of 14 questions (on Likert scale, 1 "Strongly disagree" to 7 "Strongly agree") regarding: - attitude towards the subject / statistics ("I can learn", "I will enjoy") – A1; A7; - computation and math anxiety – A2 ; A6; - beliefs regarding complexity of the subject and statistics – A3; A9; - usefulness and incidence of statistics
Beliefs and attitudes	Verify if a positive view of oneself as individual capable of statistical reasoning has been developed; willingness to think statistically;	

	appreciate the power of statistics; understand the importance of being critical about statistical messages or arguments.	in everyday life and future profession – A4 ; A5; A8; A10; - Critical attitude towards statistical messages – A11; A12; - Self-assessment of computer skills – A13; - Self-assessment of math skills – A14.

Items in Level 2 (Statistical Knowledge) and part of the items in Level 3 (L3.1, L3.3) were adapted from Watson (2003). Assessment of Levels 4 (Context Knowledge) and 5 (Critical skills) from the Knowledge component was not included in the questionnaire considering time and length restrictions.

The dispositional aspects refer mainly to the inclination and willingness to activate the five cognitive knowledge bases. Gal (2002) mentions that the three concepts (critical stance, attitudes, beliefs) are interconnected thus it is harder to describe them in a separate way. That is why, in order to assess the level of the dispositional aspects, all of them were merged into a set of 14 questions as detailed in the second part of Table 1. A similar assessment of this component was performed by Yotongyos (2014). The questions are partly adapted from the "Survey of Attitudes Towards Statistics" developed by Schau (1999)[8]. The original survey comes in two forms, of 28 and 36 items, but the questions were adapted to fit the aims of the study.

The questionnaire includes also information about: gender, high-school profile and self-assessment of computer and mathematical skills.

3.2 Sample description

The dataset consists of 30 questionnaires, filled in by the second year students of the Applied Modern Language program at the beginning of the "Introductory Statistics" course (March 2017). The analysis encompasses almost the entire target audience, since the "Applied Modern Language" program has a rather small number of candidates.

The respondents come almost entirely from urban residential areas, roughly three quarters of them are girls and most of them have graduated a Social science profile high-school, as shown in Table 2.

[8] http://www.evaluationandstatistics.com/sitebuildercontent/sitebuilderfiles/sats28pre.pdf

Table 2 – General characteristics of the studied sample; ***source: authors' calculations based on the questionnaires' results***

Gender distribution		**Residential area distribution**		**High school profile distribution**	
Girls	77%	Urban	90%	Social Sciences	83%
Boys	23%	Rural	10%	Math – Informatics	17%

Taking into consideration the small sample size, any inference or conclusions will be made with caution, keeping in mind that the main aim of this analysis is to find the particularities of the "Applied Modern Language" students when learning a first Statistics Course.

4. Results and discussion

The questionnaires' results are analysed by considering the following research questions:

i. The overall accuracy ratio attained for each level of statistical literacy. At this stage, the main focus will be on the overall accuracy ratios for each question, with special emphasis on the questions with the highest and lowest correct response rates.

ii. Distribution of score results, analyse the potential sources of variation for students' statistical literacy scores;

iii. Results regarding the beliefs and attitudes of students' towards statistics and statistical concepts;

iv. Investigation of several correlations between the score attained and attitudes or beliefs regarding mathematics, statistics.

4.1 Assessing students' statistical literacy level – strengths and weaknesses

The questionnaires' results are analysed on the three tiers as described in Table 1:

- Literacy skills
- Statistical knowledge
- Mathematical knowledge

The Literacy skills represent the ground level of statistical literacy. As detailed before, at this stage students had to answer 6 questions, three of them related to reading information from a histogram, while the other three referred to reading and interpreting a contingency

table. The difficulty of the questions increased from the first to the third item, consequently the rate of students giving the right answer declined accordingly. Non Response rate was zero for this set of questions.

As can be seen in Figure 1, the first two questions regarding the histogram obtained almost maximum correct rate, while the first item on the contingency table was the only question in the entire questionnaire with a 100% hit rate. The third and most difficult item from each set of questions was answered correctly by 70% of the students, well above a medium level.

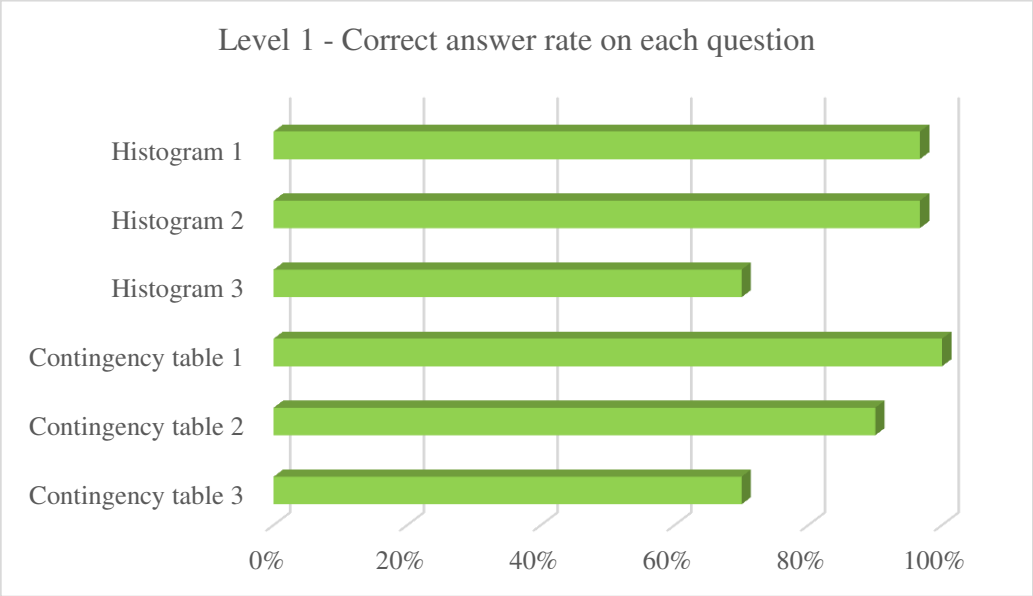

Figure 1 – Correct answer rates for the questions in the first level of statistical literacy (Literacy Skills); *source: authors' calculation based on the questionnaires' results*

Giving equal weights to these 6 items, an 87% overall accuracy ratio for the first level of statistical literacy is obtained. Thus, it can be concluded that:

- Generally, students know how to retrieve simple information from a graph and from a table;
- Students exhibit good skills when faced with making a comparative analysis based on the data from a table or synthesizing information from a simple graph (histogram or bar chart).

For the second level (Level 2 – Statistical Knowledge), students are requested an understanding of statistical concepts such as: sample, randomness, sample representativeness and bias, inference, data sources reliability, graph interpretation.

As can be seen in Figure 2, the correct answer rates vary significantly for the questions included in this level, from 7% to 73%. A question combining inference techniques and logic elements has the lowest hit rate. Only 2 students out of 30 have the correct answer for this, while the non-response rate was 27%.

In order to the overall performance at this level, different weights were applied to the questions in order to balance the competencies tested and the items' various difficulty levels. Based on these, the overall performance of the Statistical Knowledge level was calculated at 44%.

For the Statistical Knowledge level the main aspects found are:

- Students seem to have good skills for analysing graphical displays (age pyramid graph was used as item in the questionnaire)
- The basic concepts of statistics appear to be mastered superficially; confusion between sample and data obtained after the survey, between statistical units and variables, or between variables and values of the variables are relatively common;
- Generally, a bigger sample is acknowledged as a positive thing, but notions of sample representativeness and sources of bias are poorly known.
- Most unfamiliar concepts are related to the simple statistical inference elements.

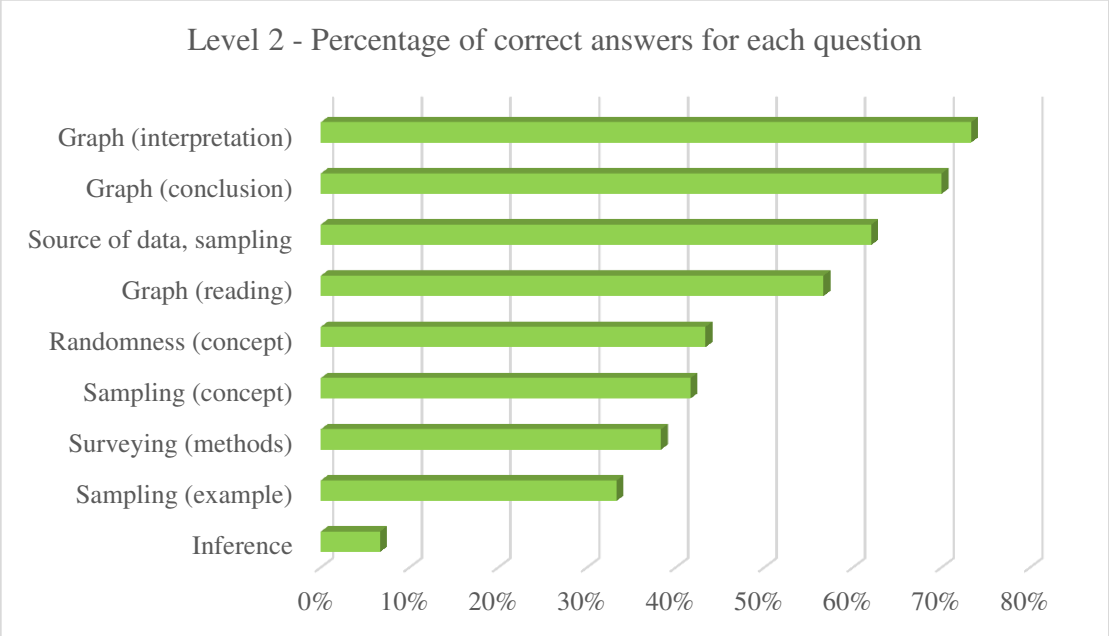

Figure 2 – Correct answer rates for the questions in the second level of statistical literacy (Statistical Knowledge); *source: authors' calculation based on questionnaires results*

On the third level of statistical literacy, the computational skills are verified (Level 3 – Mathematical knowledge). This part includes items regarding: average and median calculation and usage, percentage calculation in different contexts. The performance attained for each item included in this level are found in Figure 3.

The central tendency indicators (mean, median, mode) represent the ABC of statistics and are included in the mathematical curricula (in the 8th grade and in high-school depending on the profile[9]). It is thus expected that students know these concepts.

Overall performance on this level has been calculated at a modest 38%, based on the weights assigned for each item upon its content and difficulty.

The main conclusions that can be derived from the assessment of the Mathematical Knowledge are:

- The "median" concept, calculation and usage are poorly mastered by students; the median is correctly computed only by 17% of the students, being the question with the lowest accuracy rate from this set.
- The choice between Median and Mean (identification of main advantage of the median – not affected by outliers) is not mastered, since more than a half of respondents haven't answered this item
- Less than a half of the students can calculate correctly the simple average for a set of numbers;
- Simple relative frequencies calculation are generally known, but more complex problems involving percentages become troublesome.

[9] http://programe.ise.ro

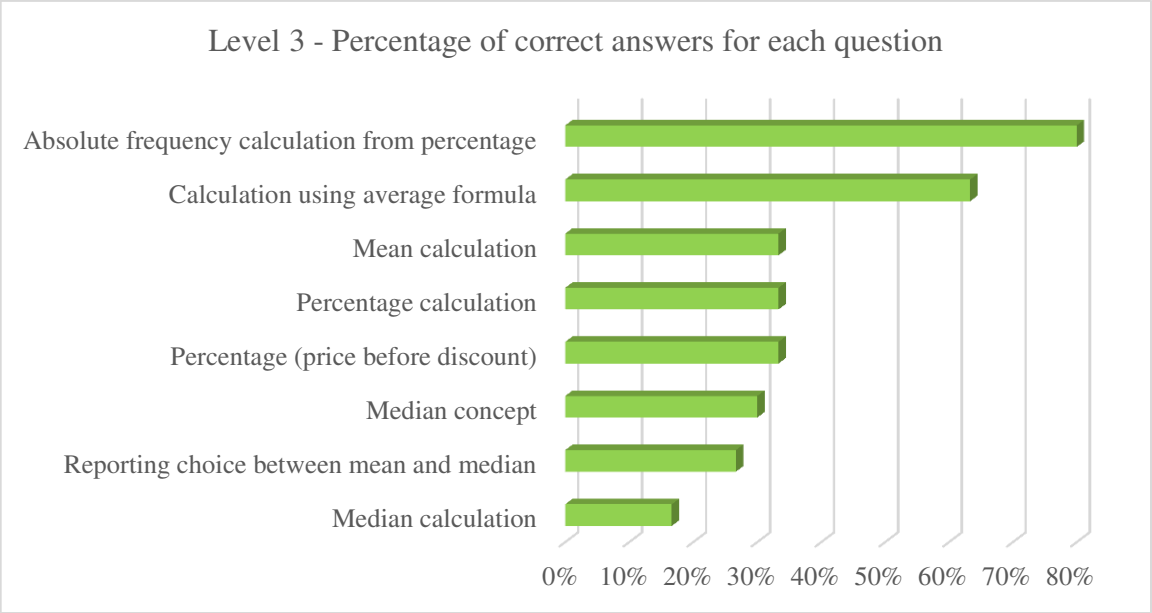

Figure 3 – Correct answer rates for the questions in the third level of statistical literacy (Mathematical Knowledge); *source: authors' calculations based on questionnaires results*

At this point, all three levels of statistical literacy have been assessed. In Table 3 below, we have summarized the overall performance rates attained at each level.

Table 3 – Synthesis of the performance rates for each level of statistical literacy; *source: authors' calculation*

Levels	Overall performance rate
Level 1 – Literacy skills	87%
Level 2 – Statistical knowledge	44%
Level 3 – Mathematical knowledge	38%

4.2 Potential sources of variation for students' level of statistical literacy

Calculating a score for each student allows determining a distribution of students' scores and evaluating potential factors triggering the differences between individual evolutions of statistical literacy skills.

In order to obtain aggregated scores at student level, each question was given a certain number of points based on the number of questions for each level and the items' difficulty level. The points given to each item are also aligned with the weights assigned for evaluating the overall performance rate at each level. In this way, the results obtained at student level will be consistent with those attained on overall level.

The total score a student can obtain is 100, divided on the three levels as follows:

- Level 1 (Literacy skills) – 24 points, split equally among the 6 items on this level;
- Level 2 (Statistical knowledge) – 40 points; 10 points given for probability notions, 16 points for sampling/ randomness related items; 6 points for the sense of data question and 8 points for interpreting the population pyramid graph;
- Level 3 (Mathematical knowledge) – 36 points; 18 points for the questions regarding averages and median; 18 points for items involving the calculation of percentages.

The students' scores range from 22 to 85.2 points, with an average of approximately 53 points. The averages scores for each levels are as introduced in Table 4.

Table 4 – Average students' scores for each level; *source: authors' calculation based on questionnaires results*

Level	**Average score**
Level 1 – Literacy Skills	21.2 points / out of 24 points
Level 2 – Statistical knowledge	17.6 points / out of 40 points
Level 3 – Mathematical knowledge	14.04 points / out of 36 points

One can notice that these are consistent with the overall performance rates introduced in Table 3 in the previous sub-section. Although it may appear that the average score is not very high, the distribution of scores (reproduced in Figure 4) shows that an encouragingly 80% of the students obtained scores higher than 40 points, while only 30% managed to get over 60 points. The most important thing to note is that half of the students are "on the average", with scores between 40 and 60 points. The scores are normally distributed as shown by the result of the Jarque – Bera test (the P-Value as high as 0.45 leads to the decision of not rejecting the null hypothesis, stating the normal distribution of the scores). It is also important to highlight the acceptable homogeneity of the dataset with a coefficient of variability of around 29% (calculated based on the descriptive statistics presented in Figure 4).

In order to determine the potential sources of variation for the students' scores, statistical tests are performed to analyse if there are significant differences for the scores obtained:

a) Between girls and boys;

b) Between students from a mathematics – informatics high-school profile and other high-school profiles.

For the first point, the null and alternative hypotheses can be defined as:

$$H_0: Average\ score\ for\ girls = Average\ score\ for\ boys$$

$$H_1: Average\ score\ for\ girls \neq Average\ score\ for\ boys$$

The t-test for the difference between two means was used and t-statistic is rather low (1,74). Although the average score calculated for girls is higher than the calculated average for boys, the sample does not supply enough evidence (P-value of 0.11) to infer the conclusion at population level. Thus, there are no significant differences between the girls' and boys' statistical literacy scores.

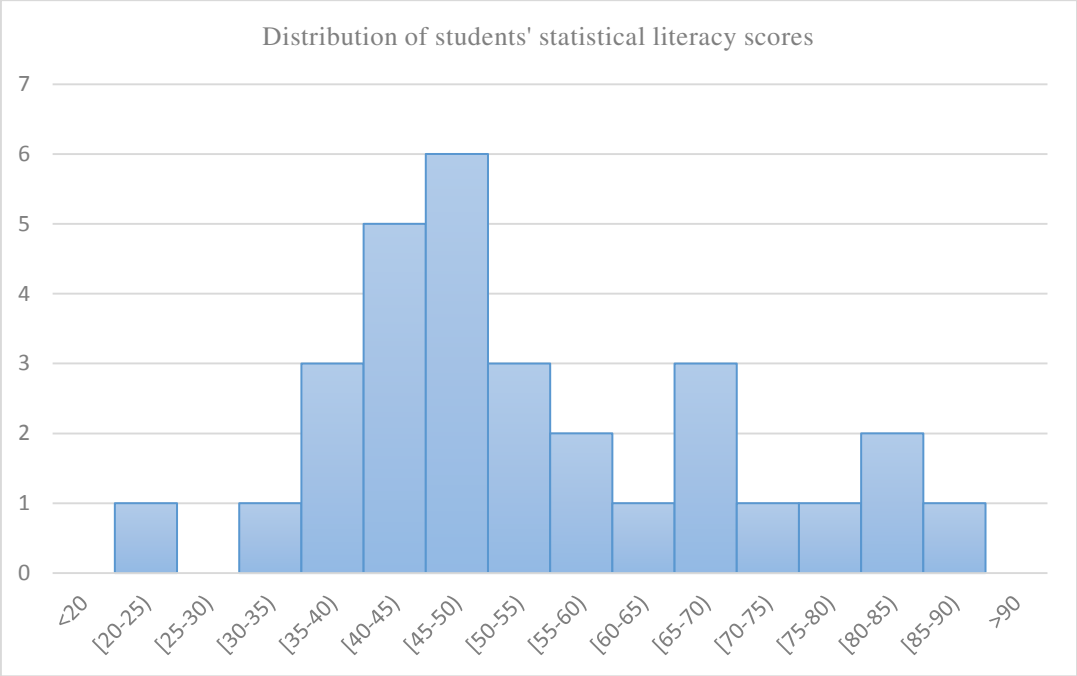

Figure 4 – Distribution of students' statistical literacy scores; *source: authors' calculation based on the questionnaires results*

The same Student test is used to identify possible differences between students from a mathematics – informatics high-school profile and other profiles and the computed value for t-statistic is 1.46. Similar to the previous case, there is not enough evidence to conclude that there are significant differences between the average score of the students with a math-informatics background and the others (P-Value is 0.18).

Both results confirm the homogeneity of the non-economists students related to their Statistical literacy.

4.3 Attitudes and beliefs towards statistics

In the questionnaire, the "Attitudes and beliefs" module was formed of 14 questions on a Likert-type scale, with possible answers from 1 ("Strongly disagree") to 7 ("Strongly agree"). For each question, the mean, median and mode have been calculated in order to evaluate an average score and identify significant variations among answers (Table 5).

Table 5 – Average results of the "Attitudes and Beliefs towards statistics" Questions; ***source: authors' calculation based on questionnaires results***

Question	Mean	Median	Mode	Average result
A1. I will enjoy the quantitative methods module	5.4	5	5	Rather agree
A2. I will feel insecure when I have to solve problems involving statistics and computations.	4.17	4	4	Neither agree, nor disagree
A3. The quantitative methods module is a complicated subject.	4.17	4	4	Neither agree, nor disagree
A4. Statistics is not useful for the profession I will embrace.	3.07	3	2	Rather disagree
A5. Statistical thinking is not applicable in everyday life.	2.67	2.5	2	Disagree
A6. The quantitative methods module involves massive computations.	4.27	4	4	Neither agree, nor disagree
A7. I can learn this subject.	5.86	6	6	Agree
A8. Statistics statements are rarely present in everyday life.	3.5	3.5	5	Rather disagree
A9. I will find it difficult to understand statistical concepts.	4.07	4	5	Neither agree, nor disagree
A10. Statistics is not relevant to my life.	2.93	3	2	Rather disagree
A11. I do not question the statistical information found in the newspaper or on the internet.	3.77	4	4	Neither agree, nor disagree
A12. I don't think that the statistical information can be misleading or biased.	3.43	4	4	Neither agree, nor disagree
A13. I have a great deal of computer experience.	4.17	4	4	Neither agree, nor disagree
A14. I am very good at mathematics.	2.83	2.5	1	Disagree

Analysing the distribution of answers for each of the questions included in this module, the following main findings are obtained:

- The positive attitude towards statistics and willingness to learn is confirmed by more than two thirds of respondents.
- Students appear to feel rather insecure when dealing with statistical calculations and consider that the subject involves computations, but not at a massive amount. We may say that the mathematics anxiety is slightly above average.

- About half of the students agree that the subject is complicated and they will find it difficult to understand the statistical concepts. In general, statistics is perceived as a complex topic.
- Two thirds of the students acknowledge the relevance of statistics, but they believe it is more relevant to their daily lives than to their profession. Although they are very aware that statistical thinking is applicable to everyday life, they don't perceive the ubiquity of statistical statements.
- 43% of students question the statistical information found in the media, while a quarter do not; however, only a promising 17% of students consider that statistical information can be misleading, putting the overall critical stance above average.
- The computer skills self-assessment reveals an average level, around 40% of respondents appreciate they have a great deal of computer experience; on the other hand, students perceive themselves with low level of mathematics: only 20% agree they are good at mathematics, but as high as 50% disagree and strongly disagree.

4.4 Correlation analysis between the Statistical literacy and the attitude towards statistics

The correlation between the attained statistical literacy performance and the attitudes towards the subject will be addressed using the scores developed at student level.

The statistical literacy students' scores haven't took into account the last 14 questions regarding the attitude and beliefs towards statistics.

The answers to these questions have been aggregated into a different score, each question being assigned the same weight. Before averaging the scores, for some questions the responses were reversed if the statement had a negatively worded. By doing this, the higher scores are assimilated with a more positive attitude. The answers have been reversed for the following questions: A2, A6, A3, A9, A4, A5, A8, A10. The reversing procedure was straightforward: the answer 1 was replaced by 7, 2 by 6, etc. Since questions 11 and 12 assess the critical stance of the students, they will be excluded from the final score. Questions 13 and 14 dealing with self-assessment of students' computer skills and mathematics will be used for correlations.

Thus, the Attitude and Beliefs score is calculated as the average of items A1 to A10 (after reversing the answer of the negative worded items as mentioned above). Higher scores mean

a more positive attitude (the maximum being 7), while lower scores are associated with a negative attitude or believe for statistics (minimum is 1).

As can be observed in Figure 5 below, the average Attitude and Beliefs is an optimistic 4.63, suggesting a slightly positive stand towards statistics. The scores range from 2.9 to 5.9, thus there are no cases with extreme negative attitude. The histogram shows that most of the students average between 4.75 and 5, confirming the positive results. The scores appear to be normally distributed (P-value of Jarque-Bera normality test is as high as 0.52).

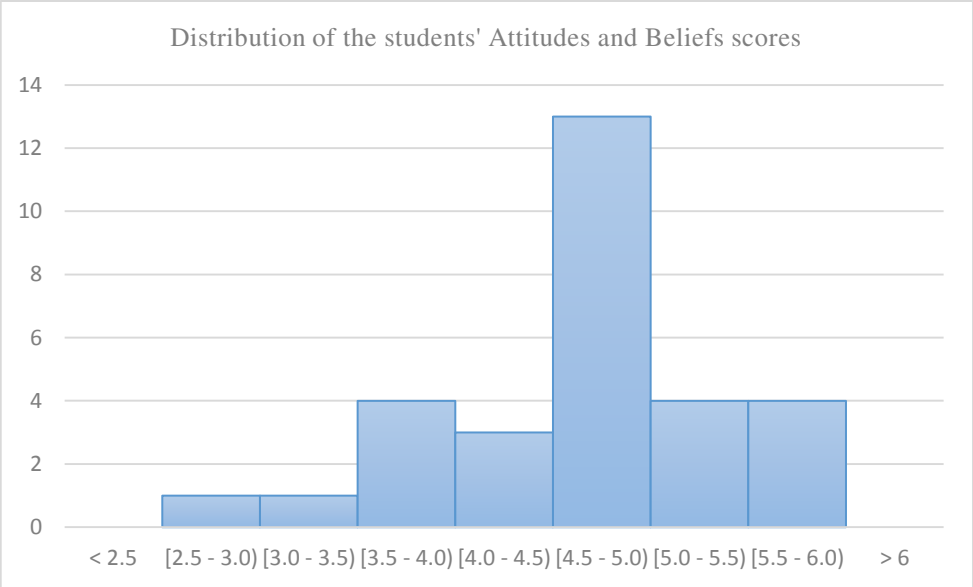

Figure 5 – Histogram of the Attitude and Beliefs scores;
source: authors' calculation based on the questionnaires results

Having both scores calculated (Statistical literacy on one hand, Attitude and Beliefs on the other hand), the correlation can be evaluated: between these scores, but also between the obtained results and students' self – assessment on mathematics or computer skills.

The assumed hypothesis is that a more positive stance (attitude) towards statistics would be associated with an enhanced perception on own mathematical and computer skills, but also with higher statistical literacy scores. The same would apply for the association between statistical literacy scores and self – assessment variables. The correlation coefficients and the associated P-values are reported in Table 6.

Table 6– Results of the correlation analysis between the attained scores and self – assessment levels; *source: authors' calculation (EViews software) based on questionnaires' results*

Variable 1	Variable 2	Correlation Coefficient	P-Value
Attitude and Beliefs score	**Self – Assessment Mathematics**	**0.3687**	**0.045**
Attitude and Beliefs score	Self – Assessed Computer Skills	-0.1773	0.35
Attitude and Beliefs score	Statistical Literacy score	0.0752	0.69
Statistical Literacy score	Self – Assessment Mathematics	0.2453	0.19
Statistical Literacy score	**Self – Assessed Computer Skills**	**0.3631**	**0.048**

Results show that only two correlation coefficients are significant (P-value under 0.05) form the five that have been tested:

- Higher Attitude and Beliefs scores are associated with a better perception of own mathematical knowledge;
- Statistical Literacy scores are moderately correlated with self – assessed computer skills;

No statistical evidence could be found regarding correlation between the Attitude and Belief level and the Statistical literacy score.

Conclusions

The starting point of this study is the acknowledgement of statistics as a daily necessity and presence in everyone's life. As most areas of knowledge, the process of teaching and learning statistics has been reshaped by the so called Data Revolution, part of the Fourth Industrial Revolution. It is this transformation of statistics' pedagogy, backed up by the motivation to give the best to our students (in terms of content, but also of methods) that trigger the unfolding of the research.

The target group of students in the current research is comprised of the Applied Modern Language programme's undergraduate students. A model was developed in order to assess their statistical literacy level and the attitudes and beliefs towards statistics.

The questionnaire analysis reveals insightful information about the characteristics of the target group. Their statistical literacy levels (graph, table reading) and critical stance are well above average and this could be further exploited in the teaching activities by including

different graphical representations and critical analysis tasks on articles found in the media involving statistical information. However, the mathematical overall level is below average (based on the items included in the questionnaire).

As the difficulty increases from one level to another in the questionnaire, the accuracy rate declines. The highest gap is between the performance registered for the literacy skills and the one for statistical knowledge, whereas one could assert that mathematical knowledge is at a small distance behind the statistical knowledge. Nonetheless, the main conclusions remain: the good literacy skills and graph interpretation that could be further used in courses, but there is a poor mastering of statistical concepts and simple statistical indicators. The analysis could not identify significant sources of variation among students' statistical literacy scores (gender and high-school profile were analysed as potential triggers).

The second part of the questionnaire dealt with the attitude and beliefs towards statistics and statistical concepts. The answers revealed a somewhat positive attitude towards statistics and willingness to learn. However, the mathematics anxiety is above average, backed by a very low self-assessment of the mathematics skills. Statistics is generally perceived as a complex subject. Nevertheless students are aware of the presence of statistics in their everyday life.

Correlation analysis revealed that a more "optimistic" attitude towards statistics is associated with a better perception of own mathematical knowledge and also that there is a moderate association between the statistical literacy level and the self – assessed computer skills.

The results offered in this paper could be further extended for other students who do not necessarily specialize in a quantitative area and thus it is expected that their mathematical skills are not highly developed. The questionnaires analysis reveals the strengths and weaknesses of the students' current statistical literacy level. Therefore, the considerable information provided by this analysis could be used to shape the teaching statistics activities for non-economists students.

References

1. Budgett S., Pfannkuch, M. (2010). Assessing students' statistical literacy. *Assessment methods in statistical education: An international perspective*, pp.103-121.

2. Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment Journal of Statistics Education 10(3), retrieved from https://ww2.amstat.org/publications/jse/v10n3/chance.html
3. delMas, R. C. (2002). Statistical Literacy, Reasoning, and Learning. *Journal of Statistics Education 10(3).* Retrieved from:

   http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html

4. Ferligoj, A. (2015). How to Improve Statistical Literacy?. *Metodoloski Zvezki*, 12(1), pp.1-10.
5. GAISE (2016), (2005). Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report. The American Statistical Association (ASA). http://www.amstat.org/education/gaise/GAISECollege.htm
6. Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International statistical review*, 70(1), pp. 1-25.
7. Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10(3), pp. 58-69.
8. Horton, N. J. (2015). Challenges and opportunities for statistics and statistical education: looking back, looking forward. *The American Statistician*, 69(2), pp. 138-145.
9. McKinsey Report (2011), Big data: The next frontier for innovation, competition, and productivity http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation
10. Ridgway, J. (2016). Implications of the Data Revolution for Statistics Education. *International Statistical Review*, 84 (3), pp. 528 – 549.
11. Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3), pp. 6-13.
12. Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: a review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, 20(2). www.amstat.org/publications/jse/v20n2/tishkovskaya.pdf
13. Watson, J. M. (1997). Assessing Statistical Thinking Using the Media, In: *The Assessment Challenge in Statistics Education*, Gal, I. and Garfield, J. B. (Eds.). Amsterdam: IOS Press and The International Statistical Institute, 107-121.
14. Watson, J. M. (2003). Statistical literacy at the school level: What should students know and do. ISI 54 Berlin 2003.

15. Watson, J., Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal* 2.2, 3-46.
16. World Economic Forum, Global Challenge Insight Report (2016). The Future of Jobs https://www.weforum.org/reports/the-future-of-jobs
17. Yotongyos, M., Traiwichitkhun, D., & Kaemkate, W. (2015). Undergraduate Students' Statistical Literacy: A Survey Study. *Procedia-Social and Behavioral Sciences*, 191, 2731-2734.
18. http://www.amstat.org/
19. http://www.worldofstatistics.org/
20. Site of the SATS (Survey of Attitudes towards Statistics): http://www.evaluationandstatistics.com/