Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v9n4p589

Promoting and assessing statistical literacy among university students. The case of Tuscany

By Valentini

Published: 15 December 2016

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 04, December 2016, 589-609
DOI: 10.1285/i20705948v9n4p589

Promoting and assessing statistical literacy among university students. The case of Tuscany

Alessandro Valentini*

Istat
Italian National Statistical Institute

Published: 15 December 2016

One of the strategic goals of Istat, Italian National Statistical Institute, is to promote statistical literacy (SL) both at central and at local level. In this context, for the first time in Italy, territorial Office of Istat in Tuscany developed a new line of research, addressed to the measurement of SL. The activity was held thanks to the cooperation with the statistical departments of the three Universities located in Tuscany (Pisa, Florence and Siena). Core of the project was the implementation of an on line questionnaire (named QValStatM) with a twofold objective. The first aim is to promote SL: the form is presented in a charming way and is associated with various statistical messages. The second aim is to measure SL in terms of number of corrected answers: each compiler immediately received his statistical profile according to the positioning of the "statistical dashboard". Over 10,000 first-year undergraduate students were given the questionnaire. Results are of a certain interest: in the range between 0 and 12 the mean score was of 7.64 which states for a hardly sufficient level of SL. As expected, results are significantly different according to various individual covariates (course attended, gender, school curricula). The present paper intends to illustrate in detail the project, the questionnaire, the method of administering, the results and implications for future strategies to promote SL.

keywords: Statistical literacy, assessment, university students, Istat

*alvalent@istat.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

Nowadays, in our increasingly hectic life context, the amount of available "information" is growing exponentially. Thanks to technological development, every single person is able to produce a "statistic" and to disclose it to an undefined audience, without any possibility of an ex-ante control. In this context of growing data deluge, measures to promote statistical literacy become much more urgent than in the past. International experiences on this topic are quite widespread. For instance in the case of medical students, MacDougall (2008) proposed ten tips for facilitating autonomous learning and effective engagement in the teaching of statistics. In order to help managers, statisticians and media relations officers in statistical organizations, Unece (2012), developed a set of guides to improve statistical literacy of different users. The approach adopted in Italy was more generalized: Istat recently constituted the "Territorial network of experts in promoting statistical literacy", a net of experts dedicated to produce materials and to plan and realize local actions for several targets, from primary schools to university (Cortese and Valentini, 2013; Valentini and Cortese, 2014). Activities of the territorial network were reinforced thanks to the full interaction between central and peripheral structures (Peci, 2013; Istat, 2015).

The scarcity of available resources required to plan the initiatives with a scientific method, and to select actions on the basis of emerging needs. This lack urgently obliged to develop instruments to measure the statistical requirements of the territories. For the first time in Italy, according to the author's knowledge, the territorial Office of Istat, in Tuscany, has taken up this challenge and has experienced system to measure statistical literacy (Valentini and Pratesi and Martelli, 2015). This new line of research, illustrated in the present paper, has been developed thanks to the cooperation with the statistical departments of the three Universities located in Tuscany (Pisa, Florence and Siena).

The driving idea of the project is to use a questionnaire (named QValStatM) with a twofold objective: to promote and to assess statistical literacy. A first prototype of the questionnaire was developed and tested in 2014 (Martelli and Pratesi and Valentini, 2014). Between February and May 2015 over 10,000 first-year undergraduate students were given a renewed version of the questionnaire: 8,600 units belonging to all the degree courses of the University of Pisa; 1,100 units belonging to Economics and Political science of the University of Florence; 700 units belonging to Economics, Business, Historical sciences and Philosophical studies of the University of Siena (Valentini and Pratesi and Martelli, 2015).

In order to capture the interest of students, the process of administering was supported in various ways: during lessons in classrooms, through news on the institutional sites of universities and so on; moreover the questionnaire is presented as if it was a game. For the occasion the slogan "if you play with statistics . . . you will see your profile!" has been coined. Once started the compilation, 12 items appear. For the most cases, items are drawn from the international context or already experimented in previous occasions and devoted to assess statistical literacy. The statistical profile assigned to each student (among the 5 available) depends on the number of correct answers, viewed through the positioning of the "statistical dashboard".

Scope of the present paper is to illustrate materials and method adopted in Tuscany to both promote and assess statistical literacy (Section 2), to show the results of the measurement (Section 3) and to design directions for future works on this topic (Section 4). More in particular, illustration of the method includes the detailed description of the questionnaire, in terms of capability of items, in order to be an instrument of measure, and of the questionnaire itself, in order to be a tool to promote statistical literacy. The process of administering and technological aspects are also treated. Results of measurement are addressed to show: the quota of corrected answers to the various items, the total score of the questionnaire (and the profile attributed) and the effect of covariates according both an univariate and a multivariate approach. Results will allow identifying the layers of students with the lowest statistical penetration. As will be shown, a great impact comes from the scholastic curricula (type of high school diploma and final evaluation), in favour of students that attended lyceum and those with the highest evaluations. Also significant is the effect of gender. These results highlight the unavoidable need to activate further initiatives to promote and support the statistical literacy through the world of school. Initiative should be devoted to students but primarily to teachers, the real channel to disseminate information into classes and in the society.

2 Materials and methods

The method adopted by Istat in Tuscany to both promote and assess statistical literacy for university students is that of an on line questionnaire, named QValStatM. Before examining details of this system, it is useful to mention that (as in Wallman, 1993), *"Statistical Literacy is the ability to understand and critically evaluate statistical results that permeate our daily lives-coupled with the ability to appreciate the contribution that statistical thinking can make in public and private, professional and personal decisions".* According to Unece (2012), this definition implies three areas *numeracy* (that is to understand quantitative data and fact), *communicating* (that is to be able to read and communicate the meaning of the data), and *discovering* (that is to discover the use of statistics for personal or professional decisions).

Actions to disseminate statistical literacy have to take into account these three aspects. Assessing statistical literacy is instead to find a concrete way to measure how a specific person is endorsed with those skills. Some of the most known projects moving towards this goal are: ARTIST (Assessment Resource Tools for Improving Statistical Thinking), a web site to assess the efficacy of statistical courses (Garfield and Gal, 1999; Garfield, 2001); a survey on students of Tasmania (Watson, 2000; Watson and Callingham, 2003); a survey on Philippines public workers (Reston, 2005; Reston, 2010) and a survey on college teachers, students and data analysts (Schield, 2001; Schield, 2006).

2.1 The questionnaire

Questionnaire QValStatM is in agreement with international literature on the subject and some items are directly drawn from the above mentioned experiences. Actual version (v. 4) of the form is the result of a 15-months experimentation process involving groups of both students and teachers. The first prototype (v. 1) included 30 items (15 close ended and 15 open ended), the second prototype (v. 2) included 15 items, most of which with close ended. For completeness even a very short 6 items version (v. 3) was experimented in the occasion of the 2014 Italian statistical days in Tuscany: results were interesting but barely discriminant, so this approach was given up.

Questionnaire is illustrated in detail in Table 1. Demands are 12, each one based on three close-ended alternative answers: one correct and two wrong. The score associated to each demand is 1 if the answer is corrected, 0 if it is wrong. Total score is the unweighted sum of scores of each item. It ranges from 0 (all answers were wrong) to 12 (all answers were corrected). Issues addressed in the questionnaire are various, logically grouped in five thematic areas: measures (how to calculate means and indicators), tables (how to read and interpret tables), figures (how to read and interpret figures); decisions (how to take decisions under uncertainty), probability (basic concepts concerning probability). The most represented area is that of figures (4 items), while all the remaining areas contain 2 items for each one.

Table 1: Questionnaire QValStatM: questions and alternative answers

Code (a)	Text and alternative answers (in bold the correct answers)
Q1 (D)	A recent newspaper article reported that in a provincial town "[...] between 2013 and 2014 thefts increased from 4 to 8, [...]". What can be assumed as trend for 2015? a. Thefts will double b. Thefts will be halved **c. Data do not allow to make any statistically valid statement**
Q2 (M)	A small object is weighed in the same balance separately from 10 students. The weights registered (in grams) are the following: 6.2; 5.8; 16; 6.2; 5.8; 5.7; 6.1; 5.9; 6.2; 6.1. If the goal is to determine as accurately as possible the weight of the object, which of the following is the most corrected method? a. Using the most common number, which is 6.2 b. Calculate the average (add ten numbers and then divide by ten) **c. Eliminate the number 16 and calculate the average of the other**

continued in the following page

Table 1 – continued from previous page

<table>
<tr><th>Code (a)</th><th>Text and alternative answers (in bold the correct answers)</th></tr>
<tr><td>Q3 (D)</td><td>Mrs. Smith wants to buy a new car, brand A or B. The objective is that the car is as resistant as possible. To take a decision, she reads specialized magazines and comes to know that on 400 cars (brand A and B), B was less resistant than A. Then she asks for advice to 4 friends of her: the first two, owner of brand B, say that their car never broke; the other two, owner of brand A, say that their car broke several times. What should be the decision of Mrs Smith, if only based on statistical information available?
a. She will buy brand B because her friends have had problems with brand A, while those who are owner of B say they have never suffered any damage
b. She will buy the brand A because the information on the magazine are based on lots of cases, and not just four
c. She can decide either for one or the other because, regardless of the brand, she can be unlucky and buy a model that needs a lot of repairs</td></tr>
<tr><td>Q4 (M)</td><td>On a newspaper article it has been written that in 50 years, in a city, the number of students has grown by 300%, from 30,000 to 90,000. Is this statement correct?
a. No, the growth is of 33.3%
b. Yes
c. No, the growth is of 200%</td></tr>
<tr><td>Q5 (T)</td><td>The table shows the labor force in Italy - December 2014 (in thousands, seasonally adjusted). Is the “percentage of employed males” equal to the “percentage of males among employed”?
<table>
<tr><th>Category</th><th>Male</th><th>Female</th><th>Total</th></tr>
<tr><td>Employed</td><td>13,071</td><td>9,351</td><td>22,422</td></tr>
<tr><td>Unemployed</td><td>1,786</td><td>1,537</td><td>3,323</td></tr>
<tr><td>Total</td><td>14,857</td><td>10,888</td><td>25,745</td></tr>
</table>
a. No
b. Yes
c. It is not possible to determine this from the table</td></tr>
</table>

continued in the following page

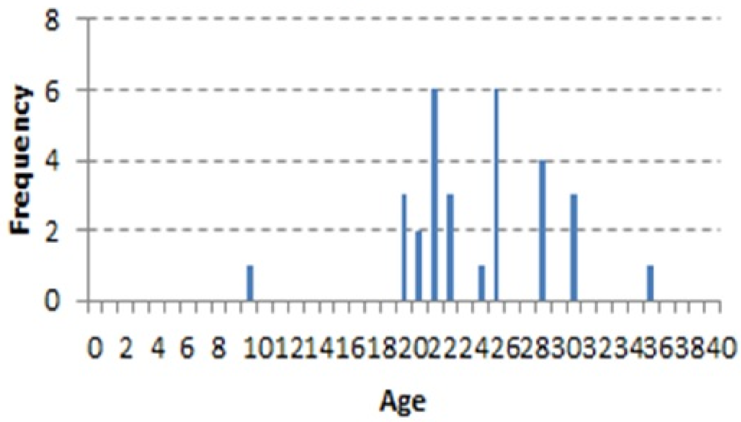
Table 1 – continued from previous page

<table>
<tr><th>Code (a)</th><th>Text and alternative answers (in bold the correct answers)</th></tr>
<tr><td>Q6 (T)</td><td>The distribution of sports played by a group of children is as follows. Comparing the two groups, can we say that tennis is more common among females than among males?
<table>
<tr><th>Gender</th><th>Football</th><th>Tennis</th><th>Basketball</th><th>Swimming</th><th>Total</th></tr>
<tr><td>Male</td><td>100</td><td>40</td><td>35</td><td>25</td><td>200</td></tr>
<tr><td>Female</td><td>5</td><td>40</td><td>40</td><td>15</td><td>100</td></tr>
</table>
a. No, because the same number of males and females play Tennis
b. Yes, because a higher percentage of females than males play Tennis
c. No, because the same number of females play both Tennis and Basketball</td></tr>
<tr><td>Q7 (P)</td><td>Two raffle boxes (A and B) contain marbles of two colors (red and black). Raffle box A has 6 red marbles and 4 black; Raffle box B 30 red balls and 20 black ones. If the goal is to draw a red ball, which raffle box has to be selected?
a. Raffle box A because it has only 10 balls
b. A or B are indifferent
c. Raffle box B, because it has 30 red marbles, while raffle box A has only 6 red marbles</td></tr>
<tr><td>Q8 (P)</td><td>At the casino a roulette player wagered 7 times on the “Red” and has always lost. Which color should he choose for the next wager?
a. The “Red”
b. The “Black”
c. It is completely indifferent</td></tr>
</table>

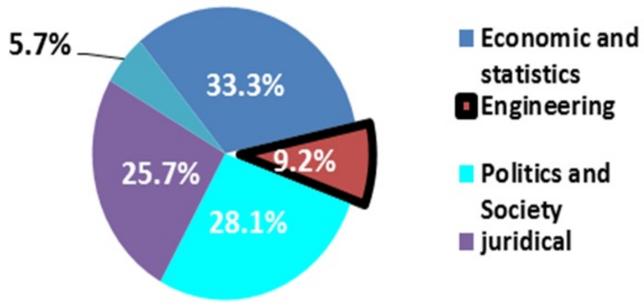
continued in the following page

Table 1 – continued from previous page

<table>
<tr><th>Code (a)</th><th>Text and alternative answers (in bold the correct answers)</th></tr>
<tr><td>Q9 (F)</td><td>The figure shows the age distribution of a group of 30 persons. What is the minimum?



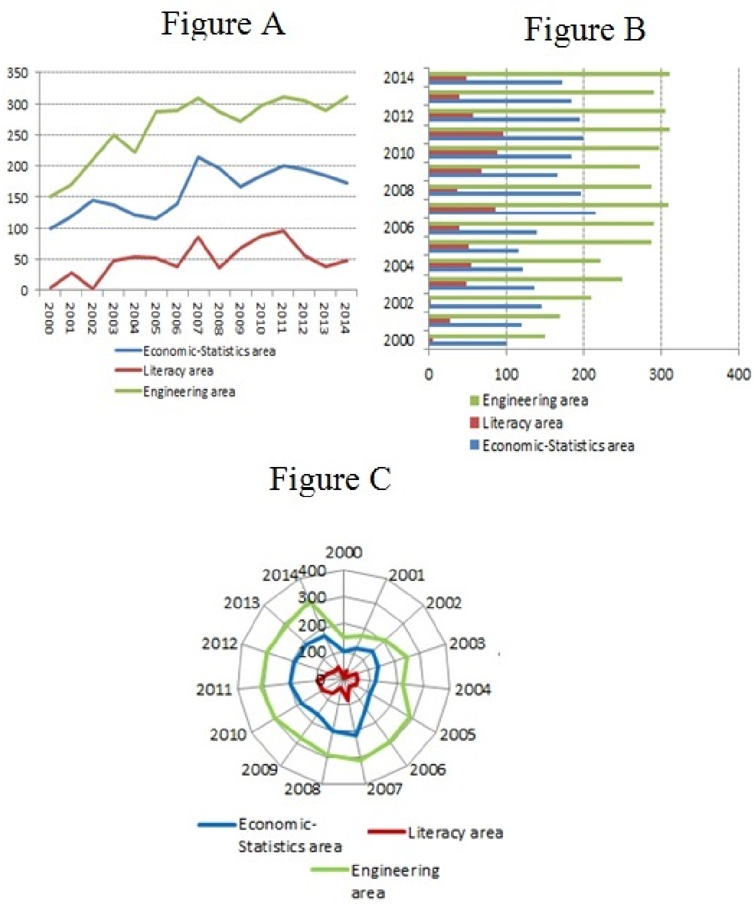
a. 1
b. Two minimum: 10 and 36
c. 10</td></tr>
<tr><td>Q10 (F)</td><td>The figure shows the distribution of students of a certain university by groups of degree courses. What does the graph intend to show?



a. "Politics and Society" group is the most numerous
b. The proportion of students in "Engineering"
c. The proportion of students in the two groups "Politics and Society" and the "Juridical" is very similar</td></tr>
<tr><td>Q11 (F)</td><td>Is there something wrong in the figure (of question Q10)
a. Yes, a label (5.7%) is outside the chart, while the others are inside
b. No, the chart has a standard format
c. Yes, the total is not exact</td></tr>
</table>

continued in the following page

Table 1 – continued from previous page

Code (a)	Text and alternative answers (in bold the correct answers)
Q12 (F)	I have the goal to use a figure to represent the time-series from 2000 to 2014 of course registrations (in thousands) in three different groups of degree courses. Three different figures (A, B and C) are available. Which is the best figure?  **a. Figure A** b. Figure B c. Figure C

(a)Areas: M = Measures; T = Tables; F = Figures; D=Decisions; P=Probability

2.2 The process of administering

The process of administering QValStatM is the core system to link dissemination actions with the measurement of statistical literacy. It is a complex operation involving at least the three main aspects illustrated below: ways of sustaining the filling of the form, solicit systems and technological aspects. As illustrated, the questionnaire contains 12 evaluative items but not only those items. It includes information concerning official statistics, the role and functions of Istat and a link to institutional web site of Italian Statistical Office. The same information was available in the email sent to all students in order to illustrate the initiative and "invite" to compile the form. It is useful to mark this aspect because dissemination activities worked for all students, even for those that decided not to compile the questionnaire.

2.2.1 How to sustain the filling of the questionnaire

The first goal of the process of administering is to maximize the response rate. At this scope, penetration of the questionnaire among students was supported in various ways: by informative email; through announcements (publication of news on the websites of the universities; affixing posters in the involved departments); by teachers (in classes during lessons); through a slogan (that allowed students to perceive the initiative as a sort of game). The slogan was "Play with statistics and ... you will discover your profile!". This refrain was probably the most effective action to persuade students to compile the form. Both the first page of the online questionnaire and the email containing the token (link to compile the form) includes a picture of the different available statistical profiles ("Seed", "Germ", "Shoot", "Tree", "Statistical tree"). The picture is reported in Figure 1. Each profile is described as follows:

- Seed: Your statistical culture needs to be improved
- Germ: You started to cultivate your statistical culture
- Shoot: Your statistical literacy is growing
- Tree: Your statistical literacy is putting down its roots!
- Statistical tree: Congratulations, you are in full flowering!

Each compiler receives a profile according to the positioning of the pointer (an angle between 0 and 180°) on the "statistical dashboard". The angle is proportional to the number of correct answers (total score), ranging from 0 to 12. In detail: if total score is 0 or 1 the profile is "Seed"; in the range between 2 and 4 the profile is "Germ"; in the range between 5 and 7, the profile is "Shoot"; in the range between 8 and 10 the profile is "Tree", if total score is 11 or 12 the profile is "Statistical tree".

Figure 2 illustrates an example of "statistical dashboard", with a number of correct answers equal to 9 (angle of 135°, profile "Tree").

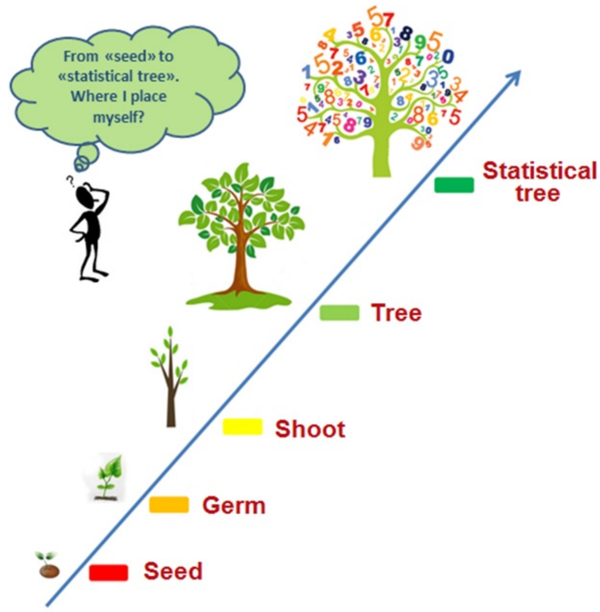


Figure 1: Picture of the different available profiles (contained in the first page of the form and in the e-mail with the invitation to compile the form)

Figure 2: The "statistical dashboard". An example associated with a total score of 9 (angle 135°; profile: "Tree")

2.2.2 Monitoring trends and responses

Data collection started the 9th of March 2015, and closed around two months later, on May, the 22nd. Around 10,000 first-year students of the three Universities of Tuscany received an email with the invitation to compile the questionnaire. The exact number of students is of 9,626:

- University of Pisa: 7,861 units, that are all students enrolled in the academic year 2014-2015, except those who have formally withdrawn before the end of the survey period (775 units). All courses degree are included. The most copious groups are those of Engineering, Economics-Statistics, Literary-Humanistic-Linguistic
- University of Florence: 1,062 units, attending the first-year of courses of Economics and Political Science
- University of Siena: 703 units, students of the first-year of Economics and Banking, Economics and Business, History of European cultural heritage, Literary and philosophical studies).

Unwilling respondent during the survey period received (almost weekly) email reminders. At the end of the survey responses were 3,071, 31.9 of units of observation. As shown in detail in Table 2, the highest response rates were recorded in Siena (37.3%), followed by Pisa (31.9%) and Florence (only 28.7%). Detailing answers for the various courses, in Siena the response rate was very good for students of Economics and Statistics (39.1%), in Florence for students of the Socio-political group (35.3%). For the students of Pisa the degree of response was more composite. The highest response rate was that of the Science group (42.5%), followed by the Geo-Biological-Chemical-Pharmaceutical group (35.6%) and by the Economics-Statistics group (35.3%). On the contrary, the lowest response rate was that of the Legal Group (26.4%) and that of the Medical group (27.1%).

2.2.3 Technological aspects

The QValStat questionnaire was implemented on line using the software Limesurvey (under GNU license). As already anticipated, each unit of the observation field received a mail containing the link (token) to compile his/her individual questionnaire. To do this the list of units and that of individual covariates (name, surname, email addresses, ID number, university, course attended, gender, type of degree, final evaluation of degree) were pre-charged on an Istat server. Responses to the questionnaire were linked to individual covariates through ID number.

During the administering process, questions do not follow the sequential order but are randomly presented with the scope to avoid cheating phenomena (Invalsi, 2011). To act the randomization, questions are divided in two parts: P1 (questions from Q1 to Q6) and P2 (questions from Q7 to Q12) and the order is casually selected from the two parts (in sequence: before the six questions of P1 and then the six questions of P2). Even the close-ended answers associated to each question (a, b, c) are administered in a random order. In this way the number of possible scheme of questionnaires is virtually infinite.

After the compliance of the last item, the on line system automatically calculates if each AQ_i item answers are corrected ($AQ_i = 1$) or wrong ($AQ_i = 0$) and calculates the total score as the unweighted sum of the evaluation to each item

$$T = \sum_{i=1}^{12} AQ_i$$

Table 2: Questionnaires sent, answers and response rates by University and by group of degree courses

University and Group of degree course	Questionnaires sent	Answers	Response rates %
University of Pisa	7,861	2,504	31.9
- Scientific Group	454	193	42.5
- Geo-Biological-Chemical-Pharmaceutical group	871	310	35.6
- Economic-Statistics group	1,120	395	35.3
- Engineering Group	1,661	541	32.6
- Agricultural-Veterinary group	438	132	30.1
- Socio-political group	502	150	29.9
- Literary-Humanistic-Linguistic group	1,284	373	29.0
- Medical group	920	249	27.1
- Legal group	611	161	26.4
University of Florence	1,062	305	28.7
- Socio-political group	300	106	35.3
- Economics-Statistics group	762	199	26.1
University of Siena	703	262	37.3
- Economics-Statistics group	516	202	39.1
- Literary-Humanistic-Linguistic group	187	60	32.1
Total	9,626	3,071	31.9

In this way it is possible to show immediately the positioning of the person on the "statistical dashboard" and to show the associated "statistical profile".

3 Results

The relevant number of answers and the high response rate make the analysis of results interesting for various aspects: correct answers to individual questions and total score, effect of covariates following both a univariate and a multivariate method, detailed study of the thematic areas of the questionnaire. Before examining results in detail, it is necessary to pay attention to the fact that, despite the high participation in the survey, it is not possible to extend the empirical evidence to the entire field of observation. In fact, the estimates may include distorting effects related to the presence of phenomena of self-selection (of unknown entity) between the group of respondents and that of non-respondents. This aspect is ignored in present paper.

3.1 Analysis of answers and total score

The first approach to results is that of the distribution of T, the total score (Figure 3). The modal value is 8 (16.5% of respondents), followed by the score 9 (14.7%) and the score 7 (13.6%). Form of the distribution is apparently normal, but the assumption of normality is statistically refused ($P < 0.0001$) because of the pronounced (negative) skewness.

According to the distribution of scores, the "statistical profile" of Tree is the most widen (43.7%), followed by the profile of Shoot (33.5%). Percentages are lower for the profiles of Statistical Tree (11.7%), Germ (10.6%) and Seed (0.5%).

A first in-depth analysis is carried out on the bases of the answers to individual questions. As illustrated in Figure 4 (left part), more than 50% of answers to three questions (Q4, Q2, Q9) are wrong. The one with the lowest rate of true answers is Q4, a simple calculation of a rate of growth ("The number of students in 50 years has grown by 300%, from 30,000 to 90,000. Is this statement correct?"). The question is simple only at first glance: in fact just 40.9% of answers were right ("No, the growth is of 200%").

A further question with a relevant percentage of mistakes was Q9 ("The figure shows the age distribution of a group of 30 persons. What is the minimum?"). Correct answer is "10", the lowest level of the age distribution, as selected by 42.3% of students. The most part of students indicated instead "1" (the lowest level of frequency and not of the distribution) or "Two minimum: 10 and 36", that are the values of ages associated with the lowest frequency.

The third question with a significant ratio of mistakes is Q2 (how to calculate the arithmetic mean in presence of outliers). In this case the answer was corrected in 48.3% of cases. The most part of students, in fact, didn't specify the need to eliminate the outlier before calculating the mean.

The quota of corrected answers to other questions ranges between 58.7% and 75.8%. An exception is Q10 (88.3%), the easiest item, which has for subject the interpretation of a figure. The breakdown of results by topic area of the questionnaire (Measures, Tables, Figures, Decisions, Probability) provides some interesting food for thought (Figure 4, right part). In a scale between 0 and 1, the mean value of corrected answers is quite different from a topic to another: the highest values are that of Decisions (0.7), Probability (0.68), Figures (again 0.68); slightly inferior the value for Tables (0.62). The lowest level is that of Measures (0.44), area that includes questions Q2 and Q4, two of the three questions with the lowest performance of respondents. A first insight into results shows that it is not correct to take for granted university students' ability for numeracy. On the contrary, the ability of reasoning resulted higher than expected. Another insight is that the ability to correctly interpret tables and graphs is not so widespread.

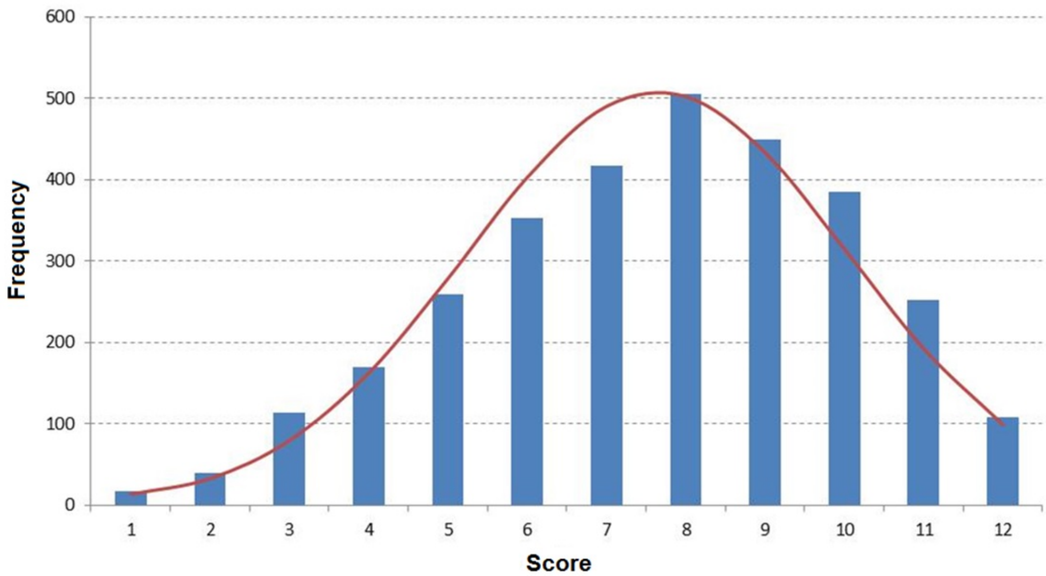


Figure 3: Frequency distribution of total score

3.2 Covariate effect

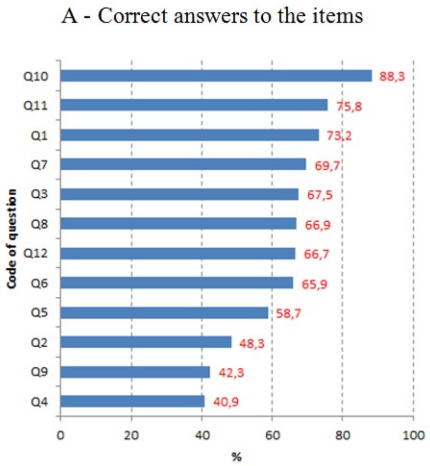
A deeper level of analysis consists in linking individual scores with covariates of target population held on administrative registers: place of origin, high school specialization, final grade and so on. Registers consists in registration files of students (held at the moment of their inscription to university). Effect of covariates is illustrated according both to a univariate and a multivariate approach. In the first case the effect of each covariate incorporates other effects; in the second case the effect is independent from other covariates.

In order to clearly show the effect of covariates, since now scores are translated in a 0 to 100 scale.

3.2.1 First-level univariate approach

A first-level analysis is realized on the basis of a univariate approach. Available covariates are six, linked to three aspect of the unit of observation, each one with two variables:

- individual characteristics: gender (male vs female) and citizenship (Italian vs foreigners)
- university choice: place of university (Florence, Pisa, Siena) and group of degree course (the list of available courses is reported in Table 2)
- previous scholastic curricula: high school specialization (Lyceum, Technical, Artistic, Pedagogical, Vocational), high school final grade, in the range from 60 to 100 (60-74; 75-89; 90-97; 98 and more).



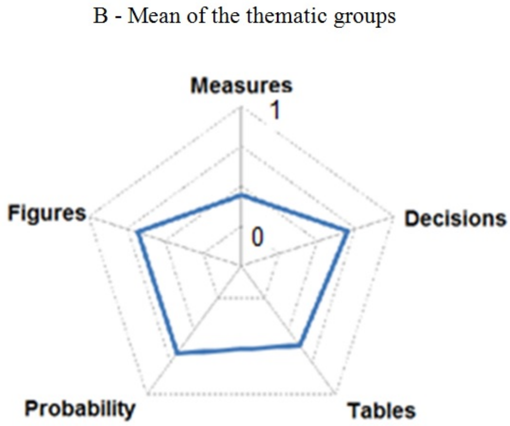


Figure 4: Distribution of correct answers to each item (left) and mean of the thematic groups (right)

As shown in Table 3, the general mean (in the range between 60 and 100) is equal to 63.7. Univariate approach for each of the six variables is statistically significant; so for each covariate the mean of each modality is statistically significant different from the mean of others. At first, let's consider individual characteristics: a clear difference by gender emerges (males: 70.3; females: 58.0); the gender gap is well known from international surveys (OECD-PISA, TIMSS) and, in Italy, is also confirmed by Invalsi tests. In the math test, males get a consistently higher score than females, a significant difference for each school level which tends to enlarge during the course of study: from 4.5 points (out of an average of 200) in primary schools to almost 10 in the upper secondary school level (Invalsi, 2014). More unusual, although somehow expected, the citizenship gap: the average level of statistical literacy for Italian students (63.8) is higher than for stranger students (58.6). Probably this is linked to difficulties with the language.

Another covariate with differential effects on the level of statistical literacy is the university choice. At first, let's consider the place of university: the average rating that emerges in the cases of Florence (64.5) and Pisa (63.9) is significantly higher than that of Siena (59.8). Both groups of the degree courses chosen have a significant impact on the measurement: students with the highest rating are those of the following groups: Scientific (78.0), Engineering (69.9) and Medical (67.5). Conversely, the lowest level was observed for students of the Legal group (53.5), of the Literary-Humanistic-Linguistic group (56.1) and of the Socio-political group (58.0).

Finally, we can state that the relationship between the level of statistical literacy and schooling is of a certain interest. Figures confirm evidence that emerged during previous tests: the strong dependency on average with the type of high school specialization and final grade. Students with an higher level of statistical literacy attended Lyceum (64.9) or a Technical school (64.0). With regard to the final evaluation, the average score is more elevated for the most talented: 72.5 students with a final evaluation of 98 and more; 65.8 for those with an evaluation from 90 to 97 and so on.

3.2.2 Second-level multivariate approach

The analysis of the differential level of statistical literacy according to the analysed individual characteristics can be further deepened through a multivariate model, which allows evaluating the main effect of each variable, net of the other, and then allowing to detect aggregation bias. The model adopted is a logit model, with the following symbolic representation.

$$logit\,[p(x)] = ln\left[\frac{p(x)}{1-p(x)}\right] = \beta_0 + \sum_{i=1}^{n} \beta_i x_i$$

The dependent variable, $logit\,[p(x)]$, is the natural logarithm of the ratio between the probability of success and that of failure to a generic item, while the independent variables (x_i) are expressed by three individual characteristics of the respondents (k= 3): gender, high school specialization and high school final grade.

Place of university, group of degree specialization and citizenship (further variables of the univariate approach) are not included in the multivariate analysis for different reasons. The place of university was excluded because personal data of students attending the University of Florence are not available. The group of degree specialization was excluded because of the strong connection, statistically significant (P <0.0001) with the high school specialization. Citizenship, instead, is not significant to a second-level analysis (this is probably due to the scarce number of cases of strangers).

To apply the logit model, the reference category for each of the three independent variables is the one with the highest level of statistical literacy in the univariate case: males ("M") for gender; students that attended "Lyceum" for the high school specialization and students with a final evaluation of "98 and more" for the high school final grade. Table 4 shows the values of the parameters of the model: exp (β) is the effect (odds ratio) arising from the transition from a reference to a no reference category for a certain variable.

The most significant effect among different variables is internal to final grade: the switch between the highest ("98 and more") and the lowest (60-74) level halves the proportion of correct answers (exp (β) = 0.53). The influence is less relevant for the other final grades. Even the effect of high school specialization is of interest: limited (-10%) the shift on odds ratio from "Lyceum" to a "Technical" school, much more significant the shift to "Artistic" school (-23%), "Pedagogic" (-24%) and in particular "Vocational" (- 34%).

4 Discussion

The present paper illustrates an experience of assessing statistical literacy conducted in Tuscany among first-year students of the three universities of the regions. The territorial office of Istat in Tuscany launched this research line in order to: understand the permeability of the statistical collective; organize appropriate feedback mechanisms; ex-post evaluates the effectiveness of initiatives. As already illustrated, in fact, assessing actions do not make any value added if kept entirely separated from those of promoting. They rather serve as a tool to address measures towards the most relevant needs, and to understand its effectiveness.

Despite the caution required in reading the outcomes of the experience (caused by the possible effects of self-selection of respondents), results clearly show that the level of statistical literacy among university students is quite “fragile”. It is sufficient a little trick on the calculation of the arithmetic mean or on the calculation of a rate to mislead the majority of respondents. This weakness has to be read at the light of the future key role that some of these young adults will cover in the forthcoming years in our society.

A deeper examination of differences in the levels of statistical literacy between the various groups of students can help us to identify some strategic actions to reduce the gap. For instance it seems of a certain interest the opportunity to carry out promoting initiatives within the school system, starting from the entrance of pupils in the education system. The implementation of promotion actions in youngsters could reduce, or stop, the gap resulting from the school performance and from the type of school that will be chosen. There is also an urgent need to pay special attention to the (growing) share of the stranger population, with regard to the activation of appropriate policies to promote, among them, the establishment of ad hoc educational packages.

More in general, in a context in which the actions of promoting statistical literacy to the school system become more and more important and prominent, National Statistical Institutes should pay specific attention to both the development of school kits and the activation of specific training courses for teachers (the most important way to direct messages to classes).

Another important project line is that to extend measurement actions to further target of particular interest. Firstly to teachers, but then also to stakeholders, for the key role that they play in society. The QValStatM questionnaire is technically ready to spread its activities to other actions, as already experienced on the occasion of the “Festival of Statistics and Demography” (Treviso, 11th - 13th September 2015). In 2016 there will be further occasions to administer the QValStatM questionnaire, for instance in the occasion of the Italian National Conference of statistics or in the occasion of the Italian statistical day. It is essential, in any case, that the actions of measurement are not conducted separately from those of promotion. Assessing without promoting is a contradiction, a sterile operation with the only purpose of detecting the existing without attempting to improve it. Also promoting without knowing statistical requirements is an inefficient operation, especially at present, in a context of scarcity of resources.

Table 3: Number of respondents, mean and standard deviation for the various covariates

Variable	Cases	Mean (range 0-10)	Std Dev	P-Sig
Total	3,071	63.7	19.9	
Gender				<0.0001
Male	*1.259*	*70.3*	*18.8*	
Female	*1.507*	*58.0*	*18.6*	
Citizenship				0.0026
Italian	*2.628*	*63.8*	*19.9*	
Stranger	*138*	*58.6*	*21.3*	
Place of university				0.0046
Florence	*305*	*64.5*	*18.9*	
Pisa	*2.504*	*63.9*	*20.1*	
Siena	*262*	*59.8*	*19.4*	
Group of degree course				<0.0001
Scientific group	*193*	*78.0*	*17.0*	
Engineering group	*541*	*69.9*	*18.3*	
Medical group	*249*	*67.5*	*19.2*	
Geo-Biological-Chemical-Pharmaceutical group	*310*	*64.5*	*18.9*	
Economical-Statistical group	*796*	*62.7*	*18.6*	
Agricultural-Veterinary group	*132*	*62.6*	*18.5*	
Socio-political group	*256*	*58.0*	*20.3*	
Literary-Humanistic-Linguistic group	*433*	*56.1*	*19.9*	
Legal group	*161*	*53.5*	*20.6*	
High school specialization				<0.0001
Lyceum	*1,649*	*64.9*	*19.8*	
Technical	*640*	*64.0*	*19.6*	
Artistic	*42*	*54.6*	*19.3*	
Pedagogic	*128*	*54.2*	*20.7*	
Vocational	*47*	*51.6*	*20.0*	
High school final grade				<0.0001
98 and more (range 60-100)	*545*	*72.5*	*18.9*	
90 97 (range 60- 100)	*351*	*65.8*	*18.1*	
75 89 (range 60- 100)	*1,018*	*62.2*	*19.8*	
60 74 (range 60- 100)	*755*	*58.8*	*20.1*	

Table 4: Parameters of the logit model

Variable	Parameters	Estimate	Effect
Gender	"F" vs "M"	-0.54	0.58
High school specialization			
	"Technical" vs "Lyceum"	-0.11	0.90
	"Artistic" vs "Lyceum"	-0.26	0.77
	"Pedagogic" vs "Lyceum"	-0.27	0.76
	"Vocational" vs "Lyceum"	-0.42	0.66
High school final grade			
	"90-97" vs "98 and more"	-0.27	0.76
	"75-89" vs "98 and more"	-0.46	0.63
	"60-74" vs "98 and more"	-0.63	0.53

References

Cortese, P.F. and Valentini, A. (2013). Cultura statistica sul territorio, nuove modalita' e nuovi attori. *NewsStat*, 9:p.13

Garfield, J. (2001). Evaluating the impact of educational reform in statistics: A survey of introductory statistics courses. *Final Report for NSF Grant REC-9732404*, 9:p.13

Garfield, J. and Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, 67: 1-12

Invalsi (2011). Esami di Stato a.s. 2009-2010. Prova nazionale, prime analisi. *Rapporto tecnico*

Invalsi (2014). Rilevazioni nazionali sugli apprendimenti 2013-2014. *Rapporto tecnico*

Istat (2015). Piano per la Promozione e lo Sviluppo della Cultura Statistica 2015. *Documento interno allegato alla deliberazione n. 81/DGEN del 14/08/2015*

Martelli, B.M., Pratesi, M. and Valentini, A. (2014). La valutazione della cultura statistica: un'indagine pilota sugli studenti dell'Universita' di Pisa *Statistica & Societa'*, 2: pp.43-48

MacDougall, M.(2008). Ten tips for Promoting Autonomous Learning and Effective Engagement in the Teaching of Statistics to Undergraduate Medical Students Involved in Short-Term Research Projects. *Journal of Applied Quantitative Methods*, 3.3: 223-204

Peci, M. (2013). La scuola superiore di statistica dell'Istat al servizio delle scuole *Induzioni*, 45 (2012): 117-120

Reston, E. (2005). Assessing Statistical Literacy in Graduate Level Statistics Education. *Paper presented at the 55th Session of the International Statistical Institute*, Sydney, Australia

Reston, E. (2010). Statistical literacy assessment and training of government personnel using data from national statistics office: Philippine context.

Schield, M. (2001). Statistical literacy: A new mission for data producers. *Statistical Journal of the IAOS*, 27, pp. 173-183

Schield, M. (2006). Statistical Literacy Survey Analysis: reading tables and graphs or rates and percentages. *International conference on teaching statistics*

Unece (2012). Making data meaningful. Part 4 : A guide to improving statistical literacy, United Nation Economic Commission for Europe. *Unece*, Geneva

Valentini, A. (2015). Come possibile misurare la cultura statistica? Resoconto di un'esperienza condotta in Toscana. *Induzioni*, 50

Valentini, A. and Cortese, P.F. (2014). Il nuovo approccio alla promozione della cultura statistica da parte della rete territoriale Istat. *Induzioni*, 48, 1: 79-93

Valentini, A., Pratesi, M., Martelli, B.M. (2015). Promozione e misurazione della cultura statistica negli Atenei della Toscana: alcune evidenze empiriche. *Statistica & Societa'*, n. 2 (2015): 36-42

Valentini, A. et al. (2014). Assessment of statistical literacy: a pilot survey on college students. The case of University of Pisa *Paper presented at the Conference of European Statistics Stakeholders*, Rome, November, 25th 2014.

Wallman, K. K. (1993). Enhancing Statistical literacy: Enriching Our Society. *Journal of the American Statistical Association*, Vol 88, no 421

Watson, J.M. (2000). Statistics in context. *Mathematics Teacher*, 93: 54-58

Watson, J.M. and Callingham R.A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2): 3-46