THE IMPACT OF COMPUTER PROGRAMS ON THE LEARNING OF DESCRIPTIVE STATISTICS: THE CASE OF DSTATS

Giuseppe Cicchitelli and Gianfranco Galmacci, Università di Perugia, Italy

The main features of the DSTATS program are presented. Remarks are also given on the results of a study conducted on undergraduate students attending a first course in statistics aimed at evaluating the effectiveness of the program for better understanding statistical concepts.

THE PROGRAM

DSTATS (Descriptive Statistical System) is a teaching tool tailored to high school- and university-level foundation course syllabuses in statistics. It covers the main areas of descriptive statistics: graphical representation, measures of location, measures of variation, measures of skewness and kurtosis, normal distribution, regression and other relationships between variables. It was created using the Visual Basic language and it runs on MS Windows platforms.

The main idea which inspired the design of the system stems from an awareness that the understanding of statistical methods can be improved by the exploration of the behaviour of the statistics concerned under different situations, real or simulated, and that this exploration can be highly fostered by the graphic features of the computer.
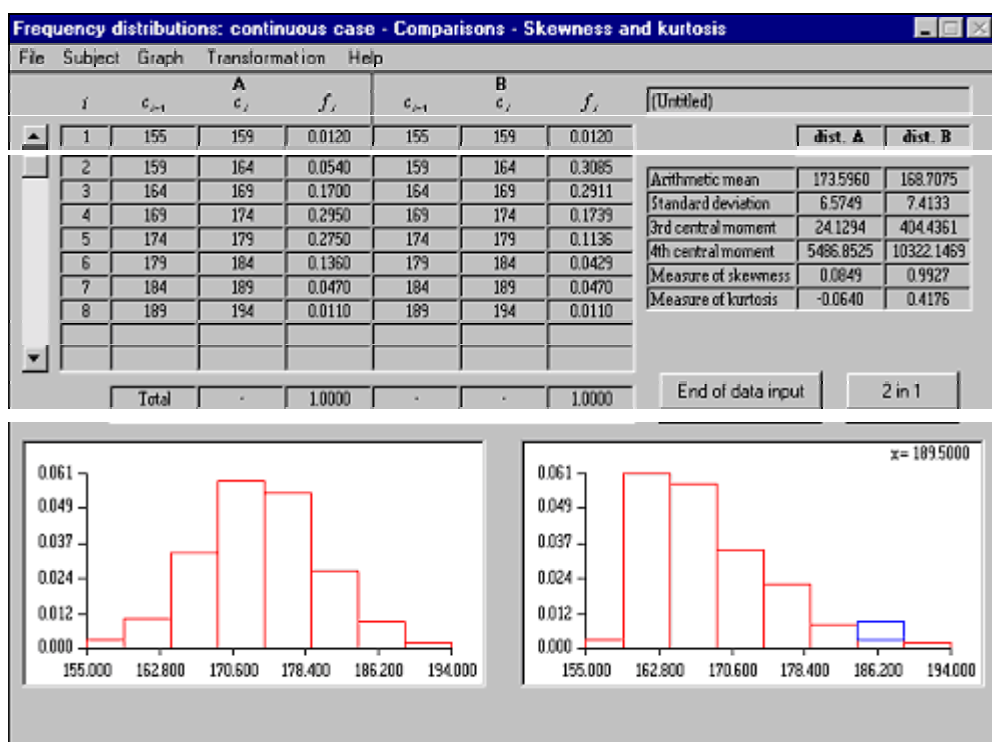
Typically, the program works from three tools: the table of "data and calculation procedures", the graphic representation, and the table of "results". A comparison is often proposed, in both tabular and graphic form, between the initial data and those modified by operating directly on the table or on the corresponding graph.

It is now worthwhile illustrating the program's architecture by presenting three specific modules as examples.

Shape of the distribution and summary statistics

The starting point is the presentation of the frequency distribution either in tabular form or graphically. Using the point-and-click system in the active window, the initial distribution can be modified. It is also possible to eliminate some classes and to add new ones. The values of the statistics selected, e.g. skewness and kurtosis measures, are updated together with the initial frequency distribution (see figure below). With this exercise, the student can see dynamically how a summary statistic behaves when

distribution characteristics are varied. This helps him to understand the nature and properties of the statistic, and also reinforces its functional interpretation.

Frequency distributions: continuous case - Comparisons - Skewness and kurtosis

File   Subject   Graph   Transformation   Help

i	A c_{i-1}	A c_i	A f_i	B c_{i-1}	B c_i	B f_i
1	155	159	0.0120	155	159	0.0120
2	159	164	0.0540	159	164	0.3085
3	164	169	0.1700	164	169	0.2911
4	169	174	0.2950	169	174	0.1739
5	174	179	0.2750	174	179	0.1136
6	179	184	0.1360	179	184	0.0429
7	184	189	0.0470	184	189	0.0470
8	189	194	0.0110	189	194	0.0110
Total			1.0000			1.0000

(Untitled)

	dist. A	dist. B
Arithmetic mean	173.5960	168.7075
Standard deviation	6.5749	7.4133
3rd central moment	24.1294	404.4361
4th central moment	5486.8525	10322.1469
Measure of skewness	0.0849	0.9927
Measure of kurtosis	-0.0640	0.4176

End of data input   2 in 1

x= 189.5000

Normal distribution

A normal distribution is shown for selected values of the arithmetic mean and variance (see figure below). The exercise consists in observing how the curve changes as the mean value and standard deviation are varied. The distribution function for a given value or, in general, the area under the curve, can be obtained by clicking on the *x*-axis. The area is highlighted in colour and the related value is shown.
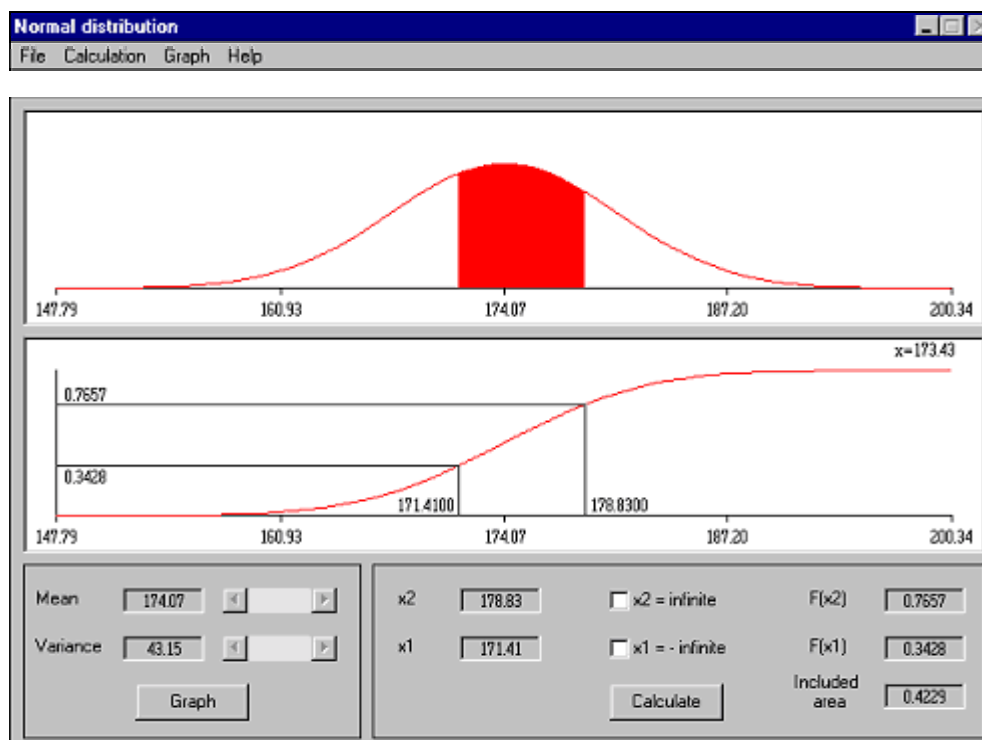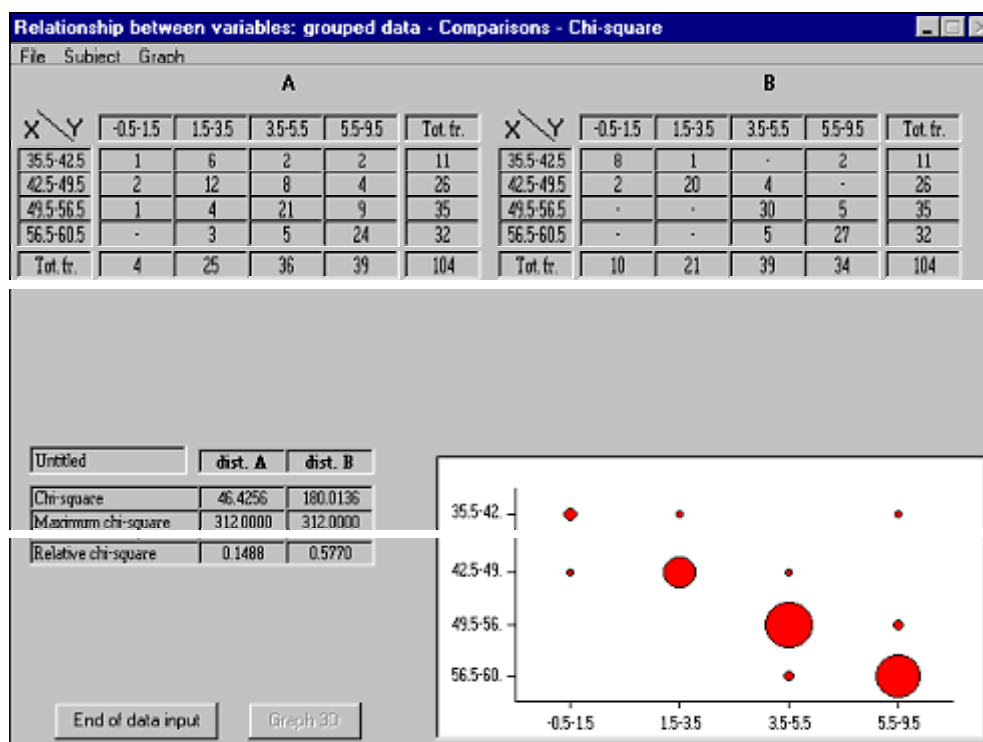
Relationship between variables (*grouped data*)

The system makes it possible to understand the link between the structure of a two-way frequency distribution, presented either in tabular form or graphically, and the value of relevant measures of relationship between two variables. Variation of the frequencies (done manually or by the point and click method) is also provided to evaluate its impact on the statistics considered (see figure below for the chi-squared statistic).

EVALUATION OF THE PROGRAM

The experimental evaluation of the DSTATS program took place within the statistics course of the Faculty of Economics at the University of Perugia. This is a second-year course, but it is taken also by students in their third year or later who are behind schedule with their exams. It is a basic, two-semester course, and in the first semester descriptive statistics is taught.

In order to evaluate the efficacy of the program, at the beginning of the current academic year, two work groups of 30 students each were formed. Chosen as relevant factors for matching were the gender, the type of secondary school attended and the preceding university career (year of enrolment and number of exams passed). It was decided to take into consideration only students who were registered for their second year in the Faculty of Economics and who had taken at least three exams - i.e., the segment of

students who were on schedule, so to speak, with their university studies. Out of this group, two random samples of size 30 were drawn, and it was verified afterwards that the two samples were homogenous in regard to the factors mentioned previously. One of the two samples was designated as the experimental group, and the other as the control group.

Relationship between variables: grouped data - Comparisons - Chi-square

File   Subject   Graph

A

X\Y	-0.5-1.5	1.5-3.5	3.5-5.5	5.5-9.5	Tot. fr.
35.5-42.5	1	6	2	2	11
42.5-49.5	2	12	8	4	26
49.5-56.5	1	4	21	9	35
56.5-60.5	·	3	5	24	32
Tot. fr.	4	25	36	39	104

B

X\Y	-0.5-1.5	1.5-3.5	3.5-5.5	5.5-9.5	Tot. fr.
35.5-42.5	8	1	·	2	11
42.5-49.5	2	20	4	·	26
49.5-56.5	·	·	30	5	35
56.5-60.5	·	·	5	27	32
Tot. fr.	10	21	39	34	104

Untitled	dist. A	dist. B
Chi-square	46.4256	180.0136
Maximum chi-square	312.0000	312.0000
Relative chi-square	0.1488	0.5770

End of data input   Graph 3D

The students in the experimental group were invited to attend laboratory exercises on the main topics of descriptive statistics, for which the DSTATS program was used. There were six exercises, for a total of approximately ten hours. The students worked in groups of two, alternating at the computer. The control group attended an equal number of exercises in the classroom on the same topics and with the same materials (data and exercises). For each exercise, the students received a written text containing the statistical data, instructions on the procedures to follow and a some diagrams, for writing down the main results.

The exercises were on the following topics: histogram and the cumulative distribution function in the discrete case and in the case of a character divided into class intervals, measures of location and quantiles, measures of variation, skewness and kurtosis indices, curve fitting and regression, relationships between variables.

The exercise was successfully completed by 49 students - 23 from the experimental group and 26 from the control group. Despite the number of students which

had dropped out, the two groups proved to be well-matched on the relevant factors mentioned above.

One fundamental problem in studies on the efficacy of the educational software is that of choosing the appropriate outcome variables (see Cohen et al., 1994, Duncan, 1993). We chose to assess the 'conceptual understanding' of the topics addressed in the experiment.

For the assessment, we devised a test composed of 85 items considered to be connected with the topics dealt with in the exercises. Students were asked to respond to each item by indicating whether it was true or false. An example of items is given below.

Consider the following two frequency distributions, A and B, and reply:

A

B

True   False
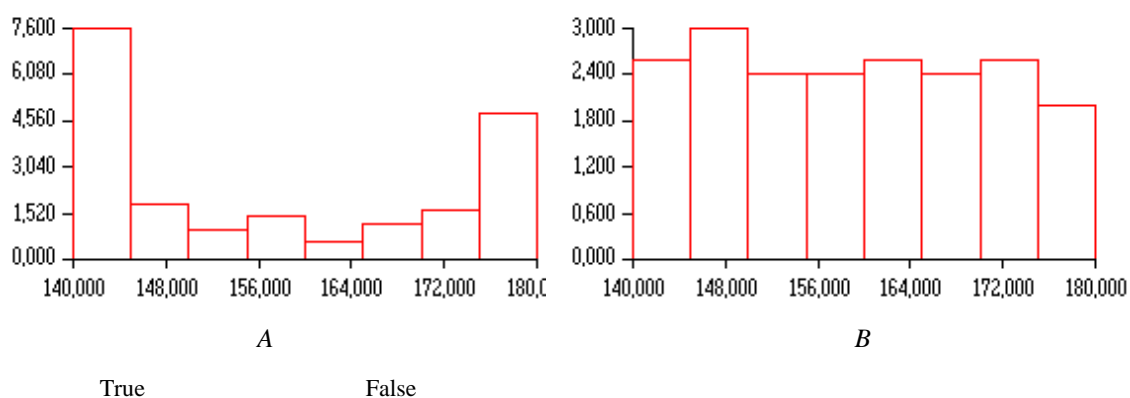
a)   The variability in *A* is lower than in *B*;

b)   The interquartile range is greater in *B* than in *A*;

c)   The concentration ratio is greater in *A* than in *B*, because the latter is nearer to the equidistribution case than the former;

d)   The difference between the quadratic mean and the arithmetic mean is greater in *A* than in *B*.

Figure 1. Example of test items.

For each test item we considered the proportion of correct responses separately for the two experimental situations. This proportion was taken as the measure of performance. We found that the experimental group did better than the control group in 70% of the cases. More refined analyses based on item response methods (see Hambleton *et al.*, 1991) to estimate the examinee's ability are underway. It would also be appropriate to model this ability, assuming as covariates students' characteristics which could affect the outcome variable.

The investigation proved useful not only for this evidence, somewhat expected, but also for the information which was collected on the difficulties students encounter in understanding statistical concepts. For example, we could observe, unexpectedly, that the test items presented in figure 1 were very difficult for both groups. The students were requested to "estimate" measures of variation from the histograms, and an implicit comparison between two different ways to express variance was proposed as well (see point d).

These findings are important because they suggest possible improvements to the software and, especially, to its application; besides, the patterns of mistakes indicate the topics which deserve more attention in the curriculum.

REFERENCES

Bartolucci F., Cicchitelli G., and Galmacci G. (Eds) (1998). *DSTATS – un ausilio informatico per l'insegnamento e l'apprendimento della statistica descrittiva*, Dipartimento di Scienze Statistiche, Univesità di Perugia, Perugia.

Cohen S, Chechile, R., Smith, G., Tsai, F. and Burns, G. (1994). A method for evaluating the effectiveness of educational software, *Behavior Research Methods, Instruments and Computers*, 26, 236-241.

Duncan, N. C. (1993). Evaluation of instructional software: *design considerations and recommendations*, *Behavior Research Methods, Instruments and Computers*, 25, 223-227.

Hambleton R.K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of item response theory*, Vol. 2, Sage publications, London.