Statistica Nova - theory and software of statistical functions mobiles

Ionel Dorofte & Ionelia Iftimie

Telmat Biz Reserch Company
Iasi,Rumanien
ionel.dorofte@gmail.com

Key Words: functions statistics mobiles, software, paradigm, psychometry

Abstract.In the last decade we presented at major international conferences a new approach to theory and calculation of empirical research. Objects that have repeatedly brought the classical theory is an insufficient adaptation to NTCI. In fact computers today, very strong, applied mostly a paradigm of classical statistical thinking ahead Computing grounded. Among others we have developed a methodology for data analysis focused on "mobile window" corollary having new theoretical concepts called new "mobile statistical functions". Basically, there are data collected through a "window" of a width sigma unit that database then scans jumping cite a measure of the variable taken into view.Se obtained in this way tens or hundreds of frequential value estimates averages of precision measurements several times better than conventional methodology. Theoretical corollary we proposed new theoretical concepts called "mobile statistical functions" of these mentioned: Frequency furniture, furniture standard deviation, coefficient of variation and correlation mobile . These functions and their immediate applications will be presented in detail in the poster will be displayed at this important congress of ISI.: We eloquent detail and practical applications for theoretical and practical advantages of this methodology. Especially important for researchers interested I made a software that allows rapid calculations and relevant views from what some reviewers of our previous work called "a new paradigm". The calculation is done to get more directly to different categories of users both in Excel Visual Basic.Net and MS.

1.Introduction. In the past five years under the name of Statistica Nova the author have suggested a series of articles to be discussed and published within important national and international scientific events focussing on the assessment of a new way of approaching the problem of analysing and interpreting empirical data in the socio-humanist sciences. This perspective has a strong branch of this demarche the assessment of the reality that the New Information and Communication Technologies (NICT) impose themselves both as a calculation and versatility power and instrumental and methodological equivalence. More than it could seem at first sight for the researcher dedicated to the daily investigation work, computer shall bring – and it already brings- new possibilities, as well as new constraints resulting in changes starting with the collection of information and documents up to analysis and conceptual patterns.

2.Historically the problem.It is obvious that we do not try to– even presumptively– to "seize" a epistemolmethodological revolution of computerizing scientific, economic and personal life in a nascendi phase which already goes on to a large extent in front of us! On the contrary due to the fact that our demarche is primarily a pragmatic one, it is also a punctual one. That is, one of the observations that we consider to have been already shown by us to the scientific community is that the algorithms that are being used by the statistic analysis programmes, however modern and sophisticated the latter ones may be, use to a large extent the calculation algorithms developed long before the current expansion of the computerised processing technologies. This discrepancy between the extensive possibilities of NTCI and classicised inertia of receptivity upon the changing of the tough nucleus of the scientific community towards this issue is obvious when we examine with a critical positive eye the statistics manuals and analysis of data coming from all socio-humanist areas. The first observation regarding the clearly dysfunctional retro style is related to the omnipresence of Greek alphabet in connection with the printing of statistic calculation formulae and coefficients difficult or practically impossible to type on computer It is difficult for us to find an explanation for this interdicted "sole" attributed to scientific players – part of a group or just in training – acting in good faith. As the observation was made public for the first time in France within the annual Conference of French statisticians we received the invitation that we shall accept, that is to prepare a small statistic calculation manual adjusted

to the interface with a personal computer.

Another remarkable inertial reminiscence is the majority procedure to entable data. It is true to call it the "romantic" stage of research in the socio-humanist sciences – as there are no technical instruments for a strict registration of massive data ad-hoc methods of data forced "grouping" in some tables with intervals in the field of considered variable had been made up. Therefore the paid sums were tabulated in groups with an opening of about 5 years, the revenue resulting from one hundred per cent, etc. It is clear that everything is done with loss of important information. As between the two extremes of the interval currently marked with one single code there is variability with possible causal determination. A large number of manuals also present sophisticated static calculation methods per grouped data (Beatrice Beaufils, 1966) which makes the calculation work extremely difficult. We should mention here that this is true for the first edition of the manual written by Yule & Kendall (1969), the first edition published in 1911 which was necessary and understandable at the time, but also for recent manuals or sites on the internet, a method similar to the way in which a car is climbed in a cart to make a complex move on a highway. Computer is more than a more advanced typing machine. It is first of all a rapid information processing tool.

As a matter of fact grouping is in tight connection with a restriction that we consider to be basic and unfortunate for the current statistic applications. When we wish to estimate the distribution values of an empirical distribution which does not have some quasi-infinity of records, we are forced to establish a scale based on a estimation interval with width h which is moved along the entire Ox sis axis and we display the value (the number of records) found in that field. Supposing we have a variable which is age and we determine a h=7 we shall note that within 0-6 years interval, we have let us say 8 fellows or a 5 % percentage out of the total number of 160. The methodology of statistics which is being currently used by most of calculation programs move this interval by one unit therefore 7-14, 15-22, 23-30, etc and calculates the number of cases presented by each grouping. From the mathematical point of view we say that we have a distribution density estimator. Histograms are immediately displayed. These ones are the most famous visualisations of data and may be rectangles, cones, circle segments or their colourful variants or in 3 D. There are several problems. The first one is related to the arbitrary determination of h interval. For different h, diagrams may have a completely different aspect. The end-user of statistic information, irrespective of his/her position- researcher or applicant, generally tends to minimize or even completely ignore these randomly controlled differences. We shall present next four diagrams of the same quasi-normal distribution (generated by a high-performance software which is but very poorly known and is free (we have not choice SPSS is king at the moment) that is "Visual Statistics". A number of four diagrams with 3, 11, 14 and 25 classes respectively have been selected. The fact that we are already aware of the fact that it is a distribution very close to the famous one -Gauss- has an impact on our perception, but the importance of the meaning may be different when the onlooker sees one of the images from a different angle. It would compulsory for the one presenting such a graphical projection to do it from several angles so that the equidistant opinion can benefit from several chances. It has always been known that lies may also be statistic. This article seeks to dismantle the methods used to do this.
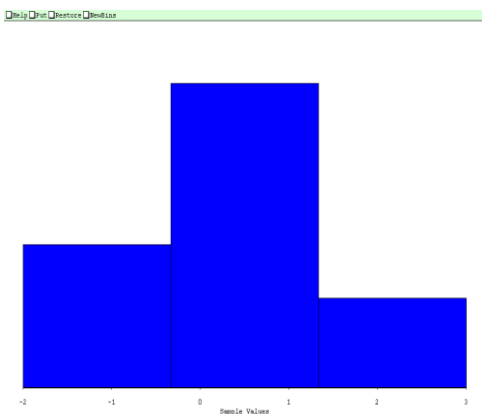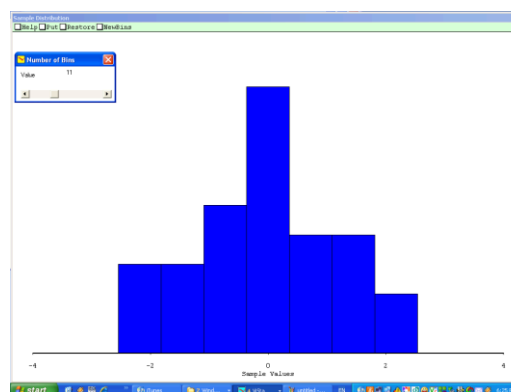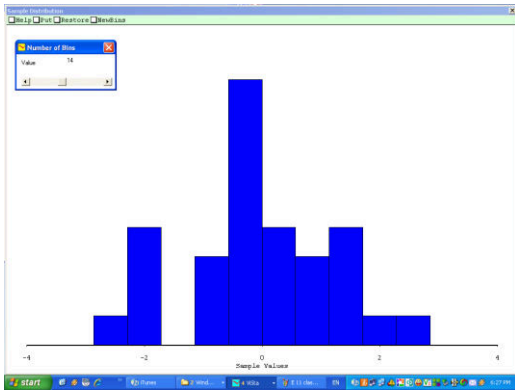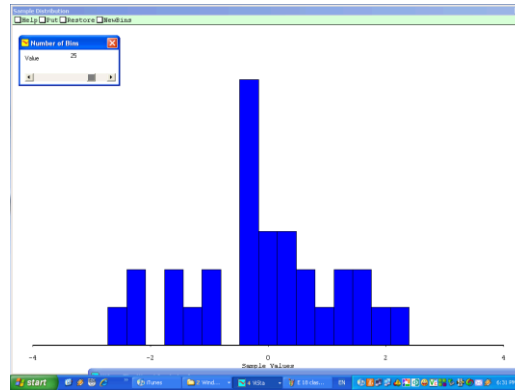

Fig. 1

Fig. 2

Fig.3

Fig. 4

Let us come back to an hypothetical example presented within the previous paragraph (with h=7 and total number of subjects equalling to 160). If we wish to control the objectivity of analysis (as a matter of fact the entire savant science of data analysis starts with these distribution density estimators) we have two possibilities: 1. We can increase h let us say to 12 or 20 or more In this case representation becomes more and more stable, but we lose any representation of the diversity of distribution on small sectors –very interesting for researchers very often.2. We can decrease h to 5, 3 or even more tempting to one single unit, but due to the fact that the group under the study is never large enough

3.Statistica Nova innovations. There shall be atypical estimates, too large differences between close cells.It is from here that started the main idea of Statistica Nova theory and software. That is we may imagine a mobile window- therefore an imaginary rectangle which scans the data moving on Ox axis with controllable h. Let us suppose for beginning that it also h=7 as within the previous example. This time the movement from left to right is not done by means of a leap by 7 units – therefore equal to h according to most statistics manuals or programs – but by a variable step l <h descending up to one single unit. This way histograms shall not have an exclusive position but crossed on the data developer. It is clear that the height of the rectangle shall be this time equal to the number of cases reported to fall within this small field. From strictly mathematic point of view, there is no problem, but we obtain a much larger number of estimates of data distribution in the \respective collection. From the conceptual point of view, we obtain a first extra inventory in the statistic theory and data analysis: mobile frequency (Ionel Dorofte.2003). Starting from here, we

have started to achieve Statistica Nova packages of programs in two versions: Visual Basic and Excel. This software allows us to launch a new important visualisation: Data curve. In the specialised or applications literature the "frequency poligon" is known as the respective

representation depends on the applicant's suppositions and options regarding the size of the estimation band. We can see next (fig. 5) such a data curve, a curve over which a "normal" curve has been concatenated-Gauss' bell as it is frequently said – with the same average and dispersion (Fig.6)
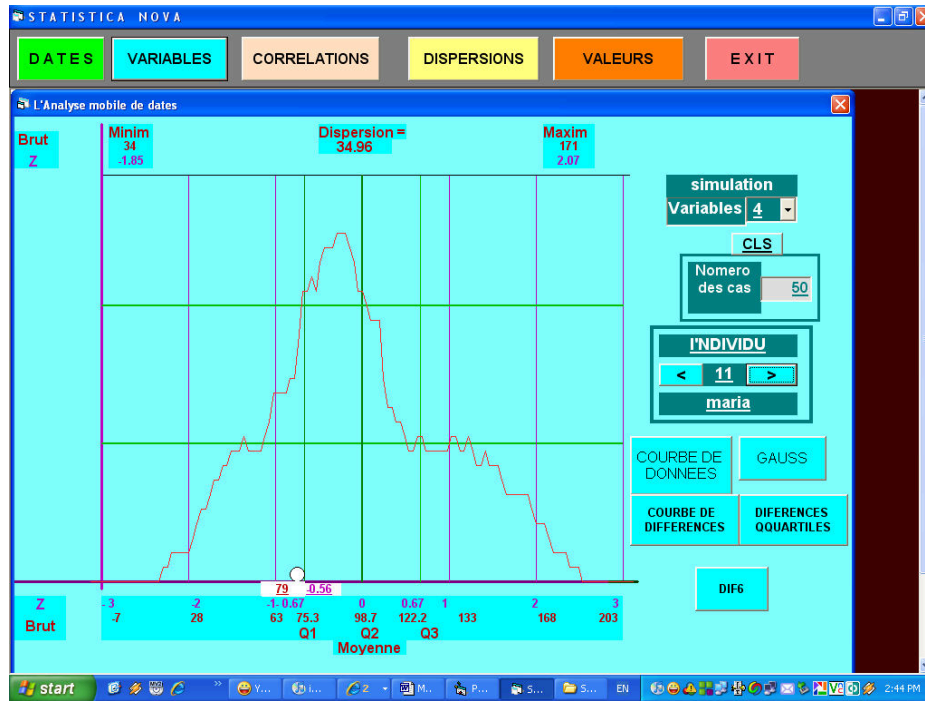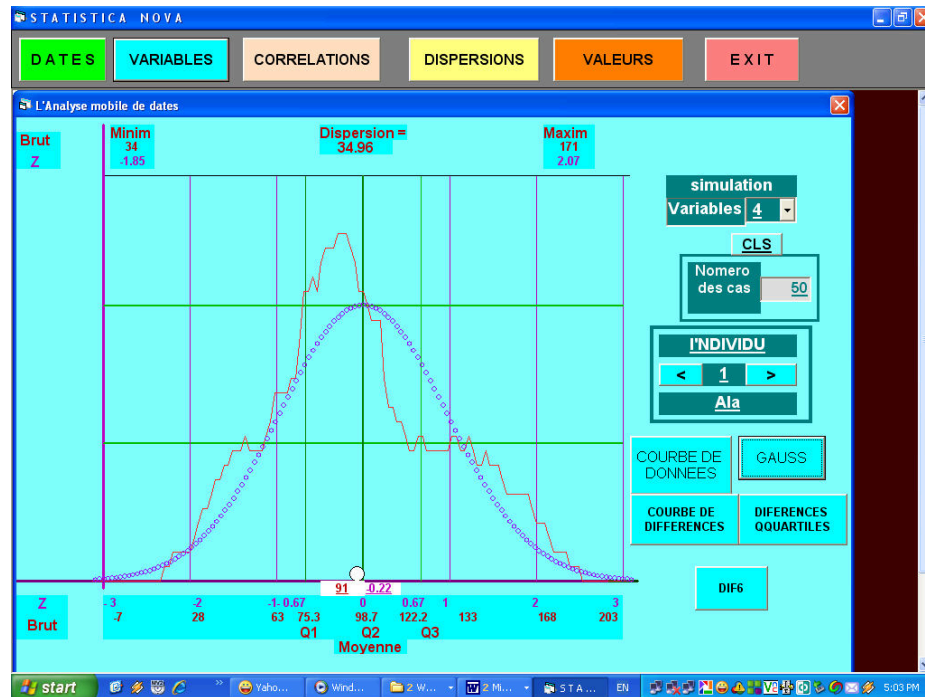
Fig. 5

Fig. 6

Even if we may not have enough space for this article, we have to notice the large number of basic information regarding the distribution of data condensed on one single "window" or image. Therefore we may read the Name of the Variable selected for presentation on the right side and that is = "simulation" and the corresponding number of the variable in the data basis "Variables = 4 "as well as the number of considered cases "Numéro des cas = 50" .

On the horizontal axis we may read on the scale whole numbers of z units starting from = "- 3" going through all values = "-2", "- 1" , "0 ", "1", "2" to the right extreme = "3". Under each value of z we may find

the gross values of the considered selection that is = "-7", "28", "63", "98.7" which is the arithmetical average, "133", "168" and "203".The researcher's or end user's work using such a visualisation is made easier by illustrating (by means of a somewhat lighter hue) I fractile, Q1= "75.3" and III fractile, Q3 = "122.2", II fractile being the average or $z = 0$ which value ="98.7" has already been presented in the above paragraph.

Gross values as well as z units are displayed so that the series of presented information can be complete in the top side of the diagram: Minimum value= "34:-1.85" and Maximum value = "171, 2.87" as well as = "34.96" dispersion.

This software is absolutely original due to the fact that it allows the graphical representation of the same diagram of general outcomes and of each individual with his/her performance which may be visually reported to the reference. With respect to figure 5 it is about the subject Maria (no. =11) with the performance = "79" in gross share or = "– 0, 56" with z units. Figure 6 with reference to the same collectivity illustrates "Individual =1; That one" with the values (91,-0.22).

Statistica Nova package of programs and adjacent theoretical system introduce some other very important concepts which are strictly connected in the new category of mobile statistic functions to be debated by the scientific community such as: mobile dispersion, mobile variant coefficient and mobile correlation coefficient, but the presentation, even if it is condensed, goes far beyond this article. Besides the punctual data about this new "paradigm" in data analysis (Ionel Dorofte 2004) the idea we would like to render is that research is being carried out with respect to the analysis of empirical data and mutatis mutandis, that is the calibration and scaling methods in general and in psychometry in particular, there are still multiple solutions which may be approached in the near future. We find this explanation to be important as the issue of measurement and metrics in psychometry is most often presented in unison by a somehow relative and easily recognizable method, the theoretical revisions and critical perspectives being no longer necessary.

4.Reference

1. **Statistica Nova : Banque des donnes et functions statistiques mobiles** (2003) Berlin, Bull.of the International Statistical Institute, vol.LX.,Book 2 (
2. **The integrated computerized system in the analysis and prediction of the human behaviour,** XXIX International Congres of Psychology, (2008), Berlin, in International Journal of Psychology, Nr.43, Issue3/4, 2008, Nr.43, Issue3/4, 2008
3. **To a revolution of mesure in psychometrics: from IQ to PQ,** *(2009), Abstracts, IMPS2009, Cambridge*
4. *Statistica Nova - la methode et le logiciel des fonctions mobiles (2010), http://hal.archives-ouvertes.fr/inria-00386763/Paris*