Statistical Software Packages as an Innovative and Motivational Tool for Teaching Statistics

David Zapletal

Institute of Mathematics and Quantitative Methods
Faculty of Economics and Admin., University of Pardubice
Pardubice, Czech Republic

Viera Pacakova

Institute of Mathematics and Quantitative Methods
Faculty of Economics and Admin., University of Pardubice
Pardubice, Czech Republic

Abstract— **Based on longtime teaching experience, the authors of the article present some reasons for the unpopularity of statistical disciplines that are taught at the economic faculties. We assume that the statistical software packages could be an effective motivational tool for teaching the statistical methods. Before explaining the individual statistical method, it is suitable to choose data sets which the students find interesting. In such a way it is possible to convince the students about the usefulness of the information provided by this statistical method. Statistical software packages allow a significant innovation of statistical disciplines. The authors present an example of good fitness tests application. The result in the form of a probabilistic model of a random variable provides detailed information of the population, much more extensive than the results of interval estimates and parametric hypotheses tests.**

Keywords— statistics; statistical software packages; motivatio; innovation; goodness of fit tests; graphic outputs

I.   INTRODUCTION

There is no doubt about the importance of statistical methods not only for economic practice but also for the whole society. The Statistics provides valuable information that cannot be obtained otherwise than by analyzing data collected from sample surveys. The demand for information, which does not consist only of statistical data but also of the results of sophisticated statistical analyses, is growing. The possibility of performing such analyzes and interest in statistics of the practice in recent decades was influenced by three factors: mass adoption of personal computers, the use of the Internet and the availability of quality statistical software packages.

Despite these facts the statistics has not been a very popular course in most universities. Introductory statistics courses are commonly considered difficult and intrinsically boring by students. Based on longtime teaching experience we will first state possible reasons for this situation and then we will suggest several ways to improve it.

Learning statistics requires students to develop a variety of skills. A successful student must gain quantitative and graphical insights along with mathematical and analytical abilities. In addition, it is sometimes difficult for students to relate statistical concepts presented in class to real world problems and everyday situations. Therefore, teaching introductory statistics requires the instructor not only to transmit knowledge, but also to enhance students' motivation and attention [1].

The basic course of statistics very often lasts only for one semester. The course usually includes the following topics: descriptive statistics, point and interval estimations, hypotheses testing, simple regression and correlation analysis, basic time series and indexes. Although the course provides frequently used basic statistical methods, it is very often boring and the students do not find it motivating. Mostly, the above suggested boredom is not caused by the topics themselves but by the way the topics are explained. This is especially true if the boards and chalk or markers are the only teaching aid used during the lectures.

As the course of the probability theory is often not taught before the course of the statistics, a lot of students do not understand such basic terms as random variable, probability distribution, distribution function etc. Students often stop following immediately at the beginning of statistical induction explanation because they do not understand the importance of such terms as random sample and sample distribution. Due to student's insufficient knowledge of probability and a low number of teaching hours, the explanation often takes the form of the instruction how to calculate this and this… It is definitely not interesting for students and it does not develop their creative thinking. Moreover, thoughtful students must feel that the seriousness of the methods is inadequate to their possible application. This feeling is also enhanced by them realizing that by using the statistical software package the results can be obtained in a few seconds.

The importance of statistical software packages in teaching and learning processes has been studied in many papers published in such journals as Journal of Statistics Education, Statistics Education Research Journal, Technology Innovations in Statistics Education, Journal on Efficiency and Responsibility in Education and Science [1, 2, 3] . The main aim of our paper is to highlight usage of statistical software to motivate students to study statistics.

Statistical software packages enable for substantial attractiveness and efficiency of education. Their use together with well prepared presentations can attract students.

Especially the first lecture ought to be motivating for students. It should focus on complex introduction of applications of statistical methods. At this stage it is ideal to used graphical outputs, which are often easy to interpret even for those who are not proficient in statistical methods. It is also necessary to choose data sets that are interesting for students. Certainly, it is not the aim of the first lecture to tell students a few definitions and terms. The most important thing is to make the majority of the students realize that the content of the lectures is interesting and useful. In our experience, the use of pivot tables in Excel is also positively motivating and their relatively simple mastering at the first seminar encourages students to further work.

To demonstrate the power, elegance, and even beauty of statistical reasoning, realistic examples from a wide variety of disciplines can persuade students that they are learning critical-thinking skills that can be applied every day and in almost any career. It is important that the motivating examples be real. Students are more easily convinced of the power of statistical reasoning if they see it applied to questions that are interesting and real to them [3].

The use of statistical software package also allows us the innovation of the subject content. As an example, goodness of fit tests that are part of virtually all statistical software packages. Various software packages differ in the number of methods for testing and, in particular, in the number of distributions for goodness of fit testing. But in our opinion, it is sufficient to make the students understand the following important fact. If we find the distribution of random variable, on the basis of random sample, we will be able to get all the information about random variable. It means all measures of the population not only of the random sample. In this case, we do not have to test or estimate anything.

Goodness of fit tests as well as maximum likelihood estimations of the parameters are mostly laborious and time-consuming [4, 5]. Thus, their wider application without using a computer and appropriate software is virtually impossible. Modeling the probability distribution of random variable is well elaborated for example in the statistical software STATGRAPHICS Centurion XV. Procedure Distribution fitting allows, on the basis of sample data, to find an appropriate theoretical model, provided such model exists. When we search for a suitable model it is possible to choose from 45 distributions (see Fig.1). The first seven distributions are discrete, others are continuous.

The output Analysis Summary of the procedure Distribution Fitting firstly provides the maximum likelihood estimates for selected type of distribution. Then we can test the correspondence of the theoretical and empirical distribution with estimated parameters. For such testing it is possible to use one of the seven of goodness of fit tests. If we find a suitable distribution the other two procedures (Tail areas and Critical values) allow us to describe the distribution in detail. The Tail areas procedure enables us to find a values of distribution
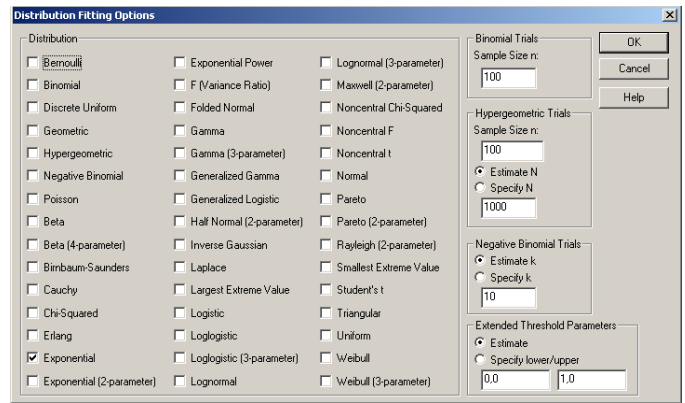

Fig. 1. Probability distributions for goodness of fit tests procedure in STATGRAPHICS Centurion XV.

function (defined as $F(x) = P(X \leq x)$) for all real values of variable x. The *Critical values* procedure allows to calculate

the quantiles. It is the inverse procedure to *Tail areas* procedure. It means the *Critical values* procedure computes the value of x for a specific value of $F(x)$, of course again defined as $F(x) = P(X \leq x)$. For example, it determines the median for $F(x) = 0,5$, upper quartile for $F(x) = 0,75$ etc.

II. RESULTS

As an example, we investigated the random variable *FNA* (Monthly expenditures of private households for food and non-alcoholic beverages in €) from the random sample of 4 705 private households in Slovakia in 2011.

The figures 2 and 3 show the graphic outputs of goodness of fit tests provided by the *Distribution Fitting* procedure from the statistical software package STATGRAPHICS *Centurion* XVI.

Data variable: FNA
4705 values ranging from 4,74 to 812,77
Fitted Distributions

Lognormal (3-Parameter)
mean = 191,91
standard deviation = 96,7712
lower threshold = -58,3228

Kolmogorov-Smirnov Test

	Lognormal (3-Parameter)
DPLUS	0,0119912
DMINUS	0,0133694
DN	0,0133694
P-Value	0,373217

Fig. 2. Parameters estimations and Kolmogorov-Smirnov test for Lognormal distribution.

We found that this random variable has a Lognormal distribution with three parameters (p-Value = 0,373217 > 0,05). The maximum likelihood estimates of parameters are included in the output in Figure 2. It means the mean $E(X) = 191,91$ € and the standard deviation $\sigma(X) = 96,77$ € are already the measures of the population – for the all private households in Slovakia in 2011, not only for the 4075 private households in random sample.

Tail Areas for FNA
Lognormal (3-Parameter) distribution

X	Lower Tail Area (<)	Upper Tail Area (>)
50,0	0,0198887	0,980111
100,0	0,149292	0,850708
200,0	0,607151	0,392849
300,0	0,874598	0,125402
400,0	0,964674	0,0353262

Critical Values for x

Lower Tail Area (<=)	Lognormal
0,05	5986,58
0,25	9870,67
0,5	13973,3
0,75	19781,1
0,99	46337,9

Fig. 3. Tail areas and critical values for fitted distribution.

If we find (as in mentioned example) the good fit of the empirical distribution with the theoretical one, all information provided by this distribution is already the information about the population not only about the sample. They are not only the measures of location or variability but also the percentiles and the distribution function values (see Figure 3). These values allow us to compute the probability of any interval of the random variable values.

Goodness of fit tests procedures always provide the graphic outputs to visual assessment of distribution fitting. The graphic outputs enable the understanding of the results of the goodness of fit tests. The histogram with fitted Lognormal distribution (Figure 5) and Q-Q plot (Figure 6) of the *FNA* confirm the result of the goodness of fit tests.
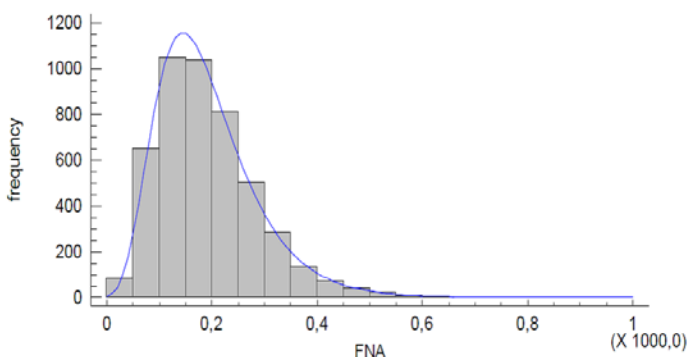

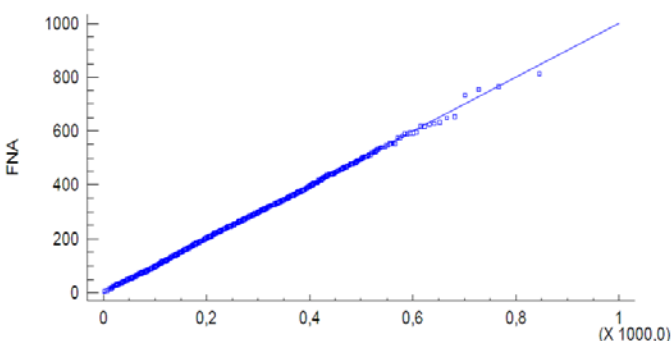Fig. 4. Histogram of fitted Lognormal distribution.

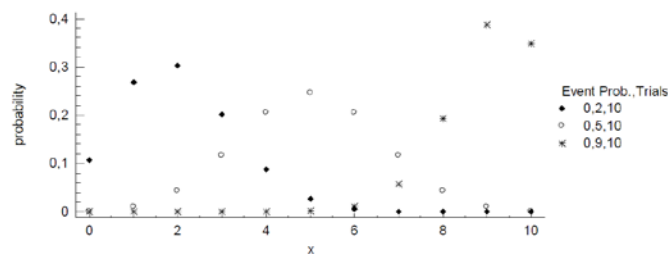Fig. 5. Q-Q plot of fitted Lognormal distribution.

Fig. 6. Binomial distribution for three different parameters (0,2; 0,5; 0,9).

Probability distribution procedure and its graphic outputs help students to understand the basic terms as probability function, probability density function, distribution function, parameters of the probability distribution and their influence on the shape of the distribution. *Probability distribution* procedure also enables students to fix the shape of individual distribution as well as the influence of changes of parameters on the shape of the distribution. As an example, the graphic output of the Binomial distribution results is mentioned (see Figure 6). Three different values of parameter p are used ($p = 0,2; 0,5; 0,9$). The use of this procedure is really effective and an illustrative way of teaching. It is apparently the only way to understand the importance and basic features of the probability distribution, especially for the students of non-mathematical faculties.

Statistical software packages are also suitable for teaching other topics as regression and correlation analysis, time series analysis etc.

III. CONCLUSION

Finally, we can say it is necessary to apply statistical methods via statistical software packages to suitable chosen real samples. To analyze real data without statistical software is practically impossible. Samples based on real data are interesting for students and they motivate them to understand the results obtained. Only such an approach can lead the students to an interest in understanding the statistical methods used.

REFERENCES

[1] J. Symanzik, N. Vukasinovic, "Teaching an Introductory Statistics Course with CyberStats, an Electronic Textbook," Journal of Statistics Education, 14, 2006.

[2] J. Poláčková, A. Jindrová, "Innovative Approach to Education and Teaching of Statistics," Journal of Efficiency and Responsibility in Education and Science, 3, 2010, pp. 14-27.

[3] G. Smith, "Learning Statistics By Doing Statistics," Journal of Statistics Education, 6, 1998.

[4] V. Pacáková, V. Labudová, L. Sipková, E. Šoltés, M.Vojtková, Štatistické metódy pre ekonómov, Bratislava: IURA EDITION, 2009.

[5] V. Pacáková, V. Labudová, L. Sipková, E. Kotlebová, Štatistická indukcia pre ekonómov, Bratislava: EKONÓM, 2012.

[6] H.H. Yang, J.C. Yu, H.J. Yang, W.H. Han, Y.J. Li, "Learning experience is important for the attitude of using statistical software," Proceedings of the 6th WSEAS International Conference on Applied Computer Science, vol. 1, pp. 166-170, 2007.

[7] A.J. Viamonte, "The Computer in the Teaching of Mathematics," Proceedings of the 6th WSEAS/IASME International Conference on Educational Technologies, vol. 1, pp. 24-29, 2010.