Journal of Information & Communication Technology
Vol. 8, No. 2, (Fall 2014) 01-06

Use & Misuse of Computer Applications In Teaching Statistics: A Case Study of Pakistan.

Hussain M A *
Institute of Business & Technology (IBT), Karachi, Pakistan.

S Murshid *
Fedral Urdu University, Karachi, Pakistan.

Safeeullah Soomro *
Indus University, Karachi, Pakistan.

ABSTRACT

We have been living in an era of modern information technology. As a result, we observe generation of voluminous amount of data in every discipline of life. To better understand the true nature of processes generating the data we need to do analysis on. Actual insights of these processes can only be understood by applying most relevant statistical techniques with the help of modern computer applications for data analysis. Correct interpretation of results obtained by these applications needs foundational know-how of both statistical subjects as well as the strengths and weaknesses of software being used for the analysis of data at hand. Such computer applications are also being used for the teaching of several computational subjects such as computational mathematics, computational statistics, computational economics, and computational finance, etc. The study aims to enhance teaching and learning of statistics by harnessing computer technology. It demonstrates the use and misuse of computer applications in teaching and research. It also emphasizes the carefulness needed by teachers, researchers and students while using such computer applications. We do comparison and contrast of a few popular commercial and open source tools (Minitab, Mathematica, Matlab, R, Statistica) that are available for faculty and students in some of the Institutes in Pakistan.

Keywords : Computer applications, Statistics, Minitab, Mathematica, Matlab, R, Statistica

---

* The material presented by the authors does not necessarily portray the viewpoint of the editors and the management of the Institute of Business and Technology (IBT) or Karachi Institute of Power Engineering.

* Husain M A : arif.biztek@gmail.com
* S Murshid : smmurshid@fuuast.com
* Safeeullah Soomro : safeesoomro@gmail.com

1. Introduction

The computing power during the past three decades has been increasing with phenomenal speed, which in turn, has resulted in the development of new methods of statistical analysis. Computers increased storage capacity and processing power, enable students to experience and explore every aspect of the statistical process. As a result, it has replaced other technological teaching aids. However, computer-based technology has also brought with it many new challenges for the teacher who seeks to determine what it has to offer and how that should be delivered to students. Several statistics packages are available (e.g., Statistica, E-views, SPSS, Minitab, etc.) varying considerably in their power and scope.

Nowadays, there are several processes generating large data sets, which need to be analyzed to get insights of these processes. Researchers need modern mathematical and statistical computer applications for data analysis. Computer applications allow students to do many things in modern ways that enhance student understanding. In some cases, it also allows students to do things that were not possible before. Computer applications are produced by using some kind of methodology, which is a systematic way of doing things. As several methodologies are available in different computer applications, we encounter different results from these applications. Moreover, correct interpretation of results obtained by these applications needs basic knowledge of statistical methods and the know-how of strengths and weaknesses of software being used for the analysis [1]. Commercial computer applications like *Statistica*, *Minitab*, *Mathematica*, *Matlab*, *R,* and *Excel* are used by researchers [2]. Researchers have shown that the reliability of statistical software cannot be taken for granted because some weak points in all random number generators can be observed. Similarly, some applications can show incorrect results for correlation analysis, one-way ANOVA output, and nonlinear least squares estimates [3].

Theoretically, there are more than hundred types of random number generators (RNGs) [4-6]. Researchers have recommended that Excel should be used for typical spreadsheet applications and for producing graphics such as bar charts, pie charts, and scatter plots but not for research level statistical analyses. Moreover, statisticians have discovered numerical instabilities in many of the algorithms present in Excel. Furthermore, the pseudorandom number generators that are used in Excel are also known to be faulty [7].

This communication has five sections. In first part above we have briefly

discussed introduction to computer-based statistics teaching. The source of data and statistical techniques are explained in Section 2. Section 3 gives some cases of statistics teaching in universities of Pakistan. In Section 4, we describe results of analysis performed in Section 3. Findings of this study and recommendations are summarized in Section 5.

2. Data and Method

At present, the use of computers in teaching of statistics is receiving increasing attention. The lack of empirical research on the role of technology in teaching statistics reveals a need for discussion on appropriate use of methodology in teaching. In this study, the data sets are generated from *Mathematica*, *Matlab*, and *R* computer applications. Ten thousand pseudo random numbers are generated from normal probability distribution to construct histograms. In the second example, cumulative probability values at different sigma () values from these applications are produced to do comparison of results.

3. Cases

This section analyzes some of the cases of difficulties in teaching of statistical concepts to students. We know that computers can be used in statistics education in several ways. It can be used as an as a computation tool, as an electronic textbook, and as a research tool. We want our students to learn statistics as potential future users, to learn how to use a professional statistical package. Although a few software had rather limited capability compared to modern statistical packages, it included necessary tools required for university level statistics course. As collecting actual data was very time-consuming, therefore, simulated datasets were used for this study. Computer technology is used to conduct simulations, which allow students to visualize and explore the long-term behavior of sample statistics under repeated sampling.

1. Students face problems in understanding and interpreting the box-plots, which are different from other statistical displays. Moreover, proper interpretation of the box-plot requires a clear concept of exploratory data analysis (EDA), which is in general, is not taught together with box-plots. Sample size is important when data do not come from a random sample.

2. The goal of software uses, however, is to use statistics to turn information into knowledge. Statistics, enhanced by technology, can

make the difference. For instance, bootstrap resampling is a simulation approach, which allows the user to get reasonable confidence intervals of estimated parameters. The accuracy of simulation method is dependent on quality of random number generator implemented in the software being used. Simulation approach is recommended in case of small sample size. The data sets used for the analysis of quality of random number are generated from *Mathematica*, *Matlab*, and *R* computer applications. It has been observed that *Mathematica* gives better result as compared to other applications.

3. Sometimes students can recognize patterns of positive and negative skew in histograms, yet are unable to translate such patterns into useful conclusions concerning the process under investigation. Students need ample knowledge of principles of good statistical graph construction.

4. Students face problem reading 3D bar charts as compared to 2D graph.

5. Students also face difficulties in applying linear models to real world problems under various conditions. Students apply data analysis approach incorrectly.

6. Random experiment, for example, tossing a coin, is explained using random numbers. It also helps explaining concept of probability.

7. We have carefully analyzed some software and the requirements of software to facilitate the teaching and learning of statistics.

Next, we obtain cumulative probability values at 4, 5, 6 locations produced from *Matlab*, *Mathematica*, *R*, *Statistica*, and *Minitab* computer applications.

Table 1 shows the cumulative probability values at 4, 5, 6 locations produced from *Matlab*, *Mathematica*, *R*, *Statistica*, and *Minitab* computer applications [8].

Cumulative probability values:

Table 1: Cumulative probability values at 4, 5, 6 locations from *Matlab*, *Mathematica*, *R*, *Statistica*, and *Minitab* computer applications.

Values	Matlab	Mathematica	R	Statistica	Minitab
4π	1.00000	0.999968	**0.9999683**	0.999968	0.99997
5π	1.00000	1.00000	**0.9999997**	1.00000	1.00000
6π	1.00000	1.00000	**1.00000**	1.00000	1.00000

4. Results and Discussions

Students have shown that histograms for 10000 data points from different computer applications are slightly different. The degree of symmetry of the histogram from the *Mathematica* seems to be good. Well, it requires further statistical test to check for degree of randomness of generated random numbers. It means that the algorithm for RNG in the *Mathematica* is one of the best algorithms. Table 1 gives cumulative probability values at 4, 5, and 6 locations from *Matlab*, *Mathematica*, *R*, *Statistica*, and *Minitab* computer applications. From this table it is obvious that *R* computer application gives better results as compared to other applications. At 6, even *R* is unable to show reasonable value of cumulative probability, which is required for six management courses. One should use advanced commands of these computer applications to get accurate results at more decimal places. We conclude this study by offering some suggestions to be addressed regarding the role of technology in statistics education.

5. Conclusion

Traditional Pakistani statistics students usually emphasize computation and neglect the development of a broader integrated view of statistical problem solving. Computer applications provide the opportunity to create new learning environment in problem solving and to foster conceptual development. The study performed reveals that almost all computer applications have limitations due to implementations of various algorithms to solve the real world problems. As a result, slightly different and sometimes, non-significant estimated parameters of the models under study are obtained. It is therefore, recommended that the researchers should be able to counter verify the critical results by using parallel applications. The most important difference between these applications is the default probability values. *Minitab* seems to be better in providing goodness of fit tests. SPSS and *Minitab* are user friendly applications. It is also recommended that ***Excel*** should be used for typical spreadsheet applications and for producing bar charts, pie charts, and scatter plots but not for research level analyses. *Mathematica* seems to be good for simulation related problems. At 6, even *R* computer application is unable to show reasonable value of cumulative probability, which is required for six management courses. In future study, more strengths and weaknesses of computer applications will be explored by taking research problems. One should use statistical techniques carefully. In this way statistics can be used to answer important questions related to decision making. This study has described how computer applications can be

used to help students become aware of their misconceptions by comparing outputs from several applications. There is a need to find effective ways to guide both teachers and research students and to provide continuing support for them as they both learn statistics using statistics software.

REFERENCES:

[1] Hussain M A and Rahim I (2012), "Teaching Optimization Techniques using *Matlab* and *Mathematica*: A Comparative Study", *Journal of Information & Communication Technology,* **6(1)**, pp. 18-26.

[2] Hussain M A, and Khan A A (2014), "Monte Carlo Simulation of Environmental and Business Data with *Matlab* and *R*", *Proceedings* 12th *International Conference on Statistical Sciences*, March 24-26, Dow University of Health Sciences, Karachi, vol. **46**. ISBN-978-969-8858-14-8, Pakistan, pp. 33-46.

[3] (Accessed 12-10-2014)

[4] http://stat.ethz.ch/R-manual/R-patched/library/base/html/Random.html

[5] G. Marsaglia and A. Zaman, *A new class of random number generators*,Annals of Applied Probability, 3 (1991), pp. 462–480.

[6] Landau D. P. and Binder K. (2009), **A Guide to: Monte Carlo Simulations in Statistical Physics 3/e**, Cambridge University Press, New York, pp. 415.

[7] Chernick M. R. (2003), Introductory Biostatistics for the Health Sciences: Modern Applications Including Bootstrap, John Wiley, New Jersey, pp. 360.

[8] Hussain M. A. et. Al., (2014), "On carefulness needed while using computer applications".

(**Submitted**)