RESEARCH PAPERS

IMPROVING QUALITY IN TEACHING STATISTICS CONCEPTS USING MODERN VISUALIZATION: THE DESIGN AND USE OF THE FLASH APPLICATION ON POCKET PCS

By

BRANDON K. VAUGHN *

PEI-YU WANG **

* Assistant Professor, The University of Texas, Austin.
** Doctoral Candidate, The University of Texas, Austin.

ABSTRACT

The emergence of technology has led to numerous changes in mathematical and statistical teaching and learning which has improved the quality of instruction and teacher/student interactions. The teaching of statistics, for example, has shifted from mathematical calculations to higher level cognitive abilities such as reasoning, interpretation, and evaluation (Ben-Zvi, 2000). Visualization is one such technological tool and can be defined as the graphical display of information. The benefit of this approach is in providing the viewer with a visual means of processing information (Segenchuk, 1997). Since the 1980's, graphical aids have been viewed as an innovative strategy in education, especially for math or science related subjects. The purpose of this study is to investigate the use of Flash applets on Pocket PCs in statistics education, specifically in the quality of instruction and teacher/student interactions. Pocket PCs and PDAs are a cheaper option than computer workstations and laptops, and easier to use in team activities. A Flash applet on sampling distributions was developed for use in a Windows Pocket PC unit, and tested using active learning activities in teaching statistical concepts. Results and suggested uses of the applet are discussed.

Keywords: Statistics Education, Improving Quality, Flash Applications.

INTRODUCTION

Purpose

Visualization can be defined as the graphical display of information. The benefit of this approach is in providing the viewer with a visual means of processing information (Segenchuk, 1997). Since the 1980's, graphical aids have been viewed as an innovative strategy in education, especially for math or science related subjects. The purpose of this study was to explore the possibility of designing and using Flash applications for Pocket PCs in statistics education. A Flash application on sampling distributions was developed for use in a Windows Pocket PC unit, and suggested learning activities in undergraduate statistics classes are also presented. Results and implications of the application are discussed.

Rationale and background

Statistics courses are a prerequisite for a wide range of programs of study at universities, especially in the area of social science. The topic of statistics cultivates student reasoning and problem-solving abilities such as comprehension, planning, execution, evaluation and interpretation (Garfield, 2002). However, many statistical topics, such as sampling distribution, can be challenging for most students to conceptually grasp. Dynamic visualization is one promising strategy to help solve students' struggles with abstract statistical concepts.

Topics in a college introductory statistics course are usually divided into three areas: descriptive statistics, probability theory, and inferential statistics (Garfield & Ahlgren, 1988). Over the past 20 years, the literature has shown a trend in students not attaining an adequate understanding of basic statistics concepts nor being able to solve applied statistical problems (Garfield & Ahlgren, 1988). Based on Garfield (1995), students' misunderstandings include a variety of fallacies: representativeness, gambler, base-rate, availability, and conjunction. Table 1 shows a summary of typical contents and students' misunderstanding in each area in a typical college-level statistics course.

Area	Typical contents	Students misunderstanding
Descriptive statistics	- Measures of central tendency - Measures of variability - Measures of position - Frequency distributions and graphs	
Probability theory	- Rules (addition, multiplication) - Independent and mutually exclusive events - Random variables - Probability distributions - The binomial distribution - The normal distribution - Sampling - Central limit theorem	- Representativeness fallacy - Gambler fallacy - Base-rare fallacy
Inferential statistics	- Estimating parameters - Testing hypotheses	- Availability fallacy - Conjunction fallacy

Table 1. Students' misunderstanding in different statistics unit

To help solve students' misunderstanding and difficulties in learning statistics, use of innovative learning strategies for statistical education have been suggested (Garfield, 1995):

Students learn by constructing knowledge.

Students learn by active involvement in learning activities.

Students learn to do well only what they practice doing.

Learning is enhanced by having students become aware of and confront their misconceptions.

Calculators and computers should be used to help students visualize and explore data, not just to follow algorithms to predetermined ends.

Students learn better if they receive consistent and helpful feedback on their performance.

The great advantage of dynamic visualization in statistics education is that each student can have full control over the properties of a population, and see for how these properties are reflected in statistics (Velleman & Moore, 1996). Students can conceptualize statistics by manipulating parameters and methods. The visualization supports enhanced accessibility of many statistical conceptions by permitting the transformation of purely symbolic presentation into spatial-geometric ones, which are easier to grasp and build cognitive models with (Ben-Zvi, 2000).

Additionally, these tools allow students to answer "what happens if this is repeated a large number of times"

through direct observation. For example, with "Rice Virtual Lab"'s sampling distribution simulation, students' understanding is developed by carrying out these repetitions, controlling parameters (such as sample size or number of repetitions), and describing and explaining the behavior they observe instead of relying exclusively on theoretical probability discussions, which can often be counterintuitive to students (delMas, Garfield & Chance, 1999).

Ben-Zvi (2000) used the term "Microworlds" as a category to describe those statistical education software which demonstrate concepts and methods by using interactive experiments, exploratory visualizations, and simulations.

Though dynamic visualization is powerful in conceptualizing abstract knowledge, it has potential problems. For example, dynamic visualization may require explicit instruction about how to use it, such as real-time display of actual use or assignments which ask students to conduct particular tasks (Velleman & Moore, 1996).

In the 1980's, Minitab was a popular tool to carry out the visualization and simulation strategy in the statistics classroom. The main reason for this was that the program was written specifically as an educational tool rather than for statisticians in industry or research. The program allowed the users to compute, explore, and simulate various topics such as distributions (Dambolena, 1986), Bayesian methods (Albert, 1993), and the central limit theorem (Gordon, 1987).

Graphing calculators have been another popular tool for visualization. Potential topics which have been supported by graphing calculators include descriptive statistics, sampling, and regression and correlation (Mittag & Taylor, 1996; Windsor, 1998). Some research findings show that students' achievement and attitude are improved by the use of graphing calculators (Mittag & Taylor, 1996).

In addition to the above two tools, the Basic computer language, the Pascal computer language, Monte Carlo techniques, and spreadsheet programs have all been used as visualization or simulation tools in statistics education. The commonality among the popular visualization tools in the 1980's is that all these tools were

designed with a text-based interface. That is, though users were allowed to interact with the visualization or simulation, the interaction approaches were very limited (e.g., keyboard-operation, instead of drag-and-drop manipulation).

The 1990's: Visualization with Graphic-based Interface

With the advancement of technology and programming languages in the 1990's, the educational visualization concentrated on more graphic-based interfaces. ELASTIC (Environment for learning Abstract Statistical Thinking) was a popular software approach during this time. This software uses interactive graphics to teach fundamental statistical concepts by coupling the power of a database management system with innovative graphing capabilities. With the mouse, users can easily create histograms, bar charts, box plots, and scatter plots, and resize or reposition any graph on the screen (Rosebery & Rubin, 1989, 1990).

In addition to this program, other popular visualization approaches of this period included graphic-based interface software based on new computer languages such as LISP and C++, which worked similarly as ELASTIC (Groeneboom et al, 1996; Marasinghe et al, 1996).

The 2000's: Visualization with internet-based Interface

JAVA applets are computer applications designed for the internet. Applets are platform-independent, meaning that they can run on any operating system that has a JAVA Virtual Machine to translate applet byte codes to appropriate platform-dependent instructions. The advantages of JAVA applets include better performance in speed, inter-operability, user interaction, portability, network computing, development, and maintenance (Kamthan, 1999). Some noted JAVA applets for statistics education are Rice Virtual Labs in statistics, Elementary Statistics JAVA applets, Statlets, StatCrunch, Statiscope, PsychStat, BusinessStat, Probability by Surprise, Web Interface for Statistics Education, CUWU Stats, PSOL, StatLab, Virtual Labs in Probability & Statistics, JavaStat, Vestac, JSci and CyberStats (Dinov, 2006).

Flash technology is the latest way to carry out educational visualization and simulation in statistics. One of the examples is ConStatS developed by the Tufts University. ConStatS consists of 12 programs, grouped into five distinct parts: representing data, probability, sampling, inference, and experiments. Pull down menus are used to access datasets, exercises, experiments, and different topics.

The emergence of new carriers: hand-held devices

The hardware on which to implement the visualization software has greatly improved in this decade. The trend of the new type of device is "palm-size but versatile". Pocket PCs, for example, are one of the products which consist of a touch screen, stylus, and several buttons for input. It measures roughly 8 cm by 12 cm and costs $150 to $700. The machines are available from a number of manufacturers (e.g., Casio, Dell, Hewlett-Packard), and all use the Microsoft Windows CE operating system. These devices have been used, for example, in biology education to show students the molecular structure through the playing of molecular visualization programs. Researchers claim that Pocket PCs are supportive for students to explore protein structures in the lab meetings, brain-storming sessions, or even over lunch break anywhere where access to a desktop computer is unavailable (Gilder, Raymer, Doom, 2001). In addition to the Pocket PCs, devices such as iPod/iPhone or Eee PC are also popular hand-held type hardware which has the potential to improve efficiency in the learning environment and increase convenience of computer-assisted instruction with minimal costs.

The handheld devices are viewed to have strong potential to improve learning; however, there are few studies addressing the educational use of them, especially in the statistics education field. The shortage of relevant studies was due to the comparatively high cost of the devices when first released, and most importantly, the lack of available software which could be used with the device. In today's world, these devices have become much cheaper, especially when compared to desktop or laptop computers. As a result, in this study, the aim was to develop an educational visualization program on the Pocket PC for examining the feasibility of mobile learning in statistics classroom.

The design of the flash application on Windows Pocket PC

For examining the possibility of using hand-held device as the visualization platform in a statistics classroom, a Flash application was developed for use in Pocket PCs. This section will describe the hardware used as well as the development of the applet.

Hardware platform and programming language selection

The study used Pocket PC's as the visualization platform. The Pocket PC had a 206 Mhz Intel StrongARM processor, a 240 x 320 backlit display with 16-bit color, and 32 megabytes of RAM. The device had a "Flash Expansion" sleeve which allowed the addition of Flash memory cards that hold anywhere from four megabytes to one gigabyte of additional memory. The Pocket PC interfaces with a regular PC via a USB or serial connection. The Pocket PC's operating system, Windows CE, is a stripped down version of Windows 2000; it retains much of the familiar interface of Windows, including the Start menu.

For the software programming language, Adobe Flash CS3 professional was used in the study to develop the visualization. Adobe Flash CS3 professional provides both a friendly interface and interactive scripting language (e.g. Action Script 2.0) for instructional designers to produce animation. The output of Adobe Flash CS3 professional is an independent executable file (.exe) or Shockwave Flash (.swf) which can be run on the windows CE operating system on Pocket PC's.

Initial development for the visualization content

The application of interest in this study covered the topic of sampling distributions. The approach in designing this applet was different from designing software for a desktop computer since the limited screen size of the Pocket PC disallowed a more complicated interface and functionalities. The first challenge was to eliminate trivial concepts in the sampling distributions unit and simplify them so that they could be put in the visualization module on Pocket Pcs.

Traditionally, the content in the sampling distribution unit includes the following:

sample proportion, spread/variability of the sampling distribution,

the usefulness of sampling distribution,

sampling distribution of a sampling mean and central limit theorem.

Restricted by the hardware, the applet designed for this study only focused on the central limit theorem part and temporarily excluded the other topics.

The central limit theorem states that the sampling distribution of the mean for any population, given an adequate sample size, will approximate a standard normal distribution. Understanding sampling distributions is essential for comprehending advanced statistical techniques, but students often fail to grasp sampling distribution concepts (Dyck & Gee, 1998; Zerbolio, 1989). The difficulty in teaching central limit theorem comes from a lack of good visualizations for the students. Teachers are left with going through problems in a "cookbook" fashion, focusing on mechanics instead of the logic of the theory itself (Garfield, 1995). Therefore, students may learn how to apply the theory in hypothesis testing by comparison of calculated values to tabled values without the understanding of the probability of sample means and the reasoning behind the test.

In order to improve the teaching of central limit theorem by use of visualization with the palm-size Pocket PC, the theorem was organized into several small content chunks, which are also the key concepts of the theorem.

The distribution shape of population: No matter which shape the population is, any arbitrary finite sampling distribution is always distributed according to a normal distribution.

The sample size: As the sample size increases, the distribution of the sample average of these random variables approaches is more closely to a normal distribution. 30 observations (samples) are sufficient to result in an approximate normal distribution.

The sample number: As the number of random samples increase, the distribution of the sample average of these random variables approaches (approximately) a normal distribution. As few as 10

observations (samples) can be sufficient to result in an approximate normal distribution under certain conditions.

Addressing the above concepts, the visualization interface on the Pocket PC was designed. The details are in the following section.

User interface design

Restricted by the 320 * 260 pixel screen, four components (including 3 menus and a set of buttons) were designed based on the above three key concepts of the central limit theorem. The details of these four components are detailed in Table 1.

Figure 1 shows the interface of the applet on the Pocket PC. In Figure 1, the population distribution is set as normal, while the sample size and sample number are all set as 20. The population consists of small pictures of people. The people are stacked in such a way to form the shape of the population. As a sample is drawn, various people fall down from the population (as shown in Figure 1). These people land in the middle area and form the particular sample histogram. The mean of this sample is computed and collected in the bottom histogram. All axes on the graphs readjust should the values go beyond the default values.

Figure 2 shows another visualization on the Pocket PC. In Figure 2, the population distribution is set as skewed. Additionally, the sample size and number of random samples are all set as 10.

In Figure 3, the population distribution is set as uniform, while the sample size and sample number are all set as 30.
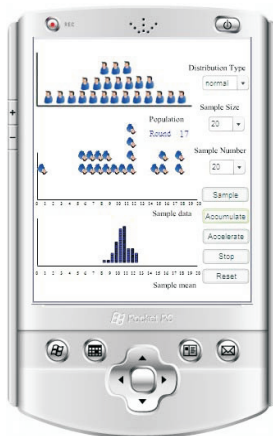
Component	Values	Design Goal
Population Distribution Menu	Skewed Distribution, Normal Distribution, Uniform Distribution	Students are allowed to see the shape of sampling distribution by selecting different shapes of the population distribution
Sample Size menu	1,5,10,20,30	Students are allowed to see the shape of sampling distribution by selecting different sample sizes.
Sample size Number	1, 5,10, 20, 30	Students are allowed to see the shape of sampling distribution by selecting different sample numbers.
Sample Button	-	Students click on the button to clear screen and initiate a new sampling procedure.
Accumulate Button	-	Students click on the button to do a accumulating sampling.
Accelerate Button	-	Students click on the button to accelerate sampling.
Stop Button	-	Students click on the button to stop sampling.
Reset Button	-	Students click on the button to clear the screen.

Table 1. The design details of the visualization program on the Pocket PC)
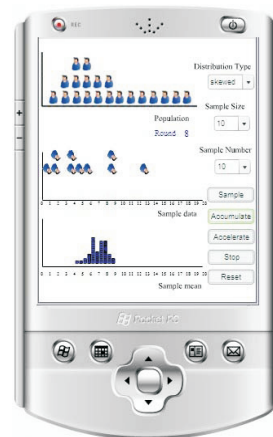

Figure 2. The visualization interface (distribution type = skewed; sample size=10; sample number=10)

Figure 1. The visualization interface (distribution type = normal; sample size=20; sample number=20)
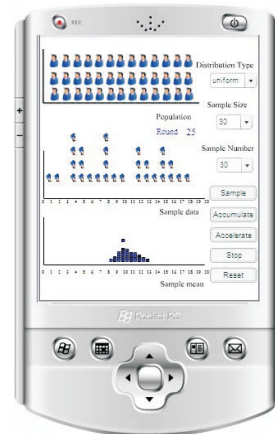

Figure 3. The visualization interface (distribution type = uniform; sample size=30; sample number=30)

As the three examples show, this visualization allows students to experience the sampling procedure as well as the central limit theorem through the interaction with the menus and buttons. Since the visualization is run on the Pocket PC, it is particularly convenient and controllable for the instructors to use in the classroom or lab settings.

Pilot Testing and Research

This applet was shown to various students in an Educational Psychology program who gave helpful comments on various changes to the applet. Additionally, a type of content validity was considered by having instructors in Learning and Cognition, as well as Statistics, look over the applet. Their suggestions were incorporated in various changes to the applet. The applet is currently being considered in empirical research comparing it to traditional computer visualizations as well as a control group.

Implementing Windows Pocket PCs into an introductory classroom

In this section, some ideas are given to show how this hand-held visualization discussed in the previous section can be incorporated into a introductory statistics course. For courses with only a short amount of time available to be devoted to this topic, engaging students in individually manipulating the hand-held visualization in the classroom allows students to better visualize the relationships between sample size, sample number, and distribution shape of the population and the sample. This could be done in about 20-30 minutes. Instructors could develop worksheets for students to complete, therefore making sure students are successfully grasping the three key ideas which the hand-held visualization aims to deliver. Having the students work with the visualization themselves is expected to have dramatic impacts on helping them to reinforce the concepts and to better prepare them to be able to solve problems that they are likely to encounter after completing the introductory statistics course.

For courses with larger amounts of time available to be devoted to this topic, group/team work is suggested for the use of the Pocket PC. Instructors might divide students into groups, and ask them to discuss the sampling results after their individual manipulation of the visualization. The students can be expected to socially construct knowledge about the central limit theorem if they are arranged to work together with the visualization program.

The affordance and constraint of Pocket Pcs

The main benefit of the Pocket PC is its portability and mobility. From the experience of designing visualization software and instructions on the Pocket PC, the researchers conclude that hand-held devices create more opportunities for students to manipulate the educational visualization anytime anywhere (as compared to traditional means such as desktop/laptop computers). This conclusion needs further justification through empirical research, which is currently being done by the researchers. In addition, with this technology, it is possible for instructors to develop applications on the device for classroom use. However, the constraints for the Pocket PC include its limited screen, smaller memory size, and longer processing/response time. It is still very challenging to organize all content chunks on the small device screen as an application and design it as an efficient instructional unit.

Conclusions

In this study, we first considered a literature review on the revolution of educational visualization in statistics. The coverage of the review was from the 1980's textual interface programs to the 2000's modern visualizations on hand-held devices. Then detail was given on the development of a visualization program on the Pocket PC. The design ideas, purpose, program interface and suggested classroom activities were discussed in the paper. This research will hopefully encourage other statistics researchers and educators to investigate and invest in mobile learning to increase the quality of students' conceptual understanding of statistics (as well as other areas of learning). Future research will consider empirical designs to examine the usefulness and efficiency of the visualization on the Pocket PC and provide more sufficient evidence of the improvement on conceptual knowledge from using the new device.

Implications

A possible implication of the study is some justification for alternative types of statistical instruction, particularly with new visualization technologies. While a radical transformation has occurred in many undergraduate statistics classes in the country, the use of handheld devices could help to alleviate the problems that computer workstations or laptops have proven in their incorporation in actual instruction (e.g., cost, battery life, bulk, and so on). By showing a possible solution of the devices, some justification can be made for instructors or schools to consider their use in instruction.

References

[1]. Albert, J. H. (1993). Teaching Bayesian Statistics Using Sampling Methods and MINITAB. *The American Statistician*, *47*(3), 182-191.

[2]. Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical thinking and learning, 2* (1&2), 127-155.

[3]. Dambolena, I.G. (1986). Using Simulation in Statistics Courses. *Collegiate Microcomputer, 4*(4), 339-344.

[4]. DelMas R.C., Garfielf, J. & Chance, B.L. (1999). A Model of Classtoom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning. *Journal of Statistics Education,* 7(3). Retrieved June 10, 2008 from www.amstat.org/publications/jse/secure/v7n3/delmas.cfm

[5]. Dinov, I.D. (2006). SOCR: Statistics Online Computational Resource. *Online Submission*, 1-22.

[6]. Dyck, J. L., and Gee, N. A. (1998). A Sweet Way to Teach Students About the Sampling Distribution of the Mean. *Teaching of Psychology*, 25(3), 192-195.

[7]. Garfield, J. (1995). How students learn statistics. *International Statistics Review, 63*(1), 25-34.

[8]. Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education,* 10(2). Retrieved June 10, 2008 from www.amstat.org/publications/j se/v10n3/garfield.h

[9]. Garfield, J. & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research. *Journal for Research in Mathematics Education, 19*(1). 44-63.

[10]. Gilder, J. R., Raymer, M. & Doom, T. (2001). PocketMol: A Molecular Visualization Tool for the Pocket PC. Bioinformatics and Bioengineering Conference, *Proceedings of the IEEE 2nd International Symposium on Bioinformatics for Drug-Development,* 11-14.

[11]. Gordon, F. (1987). Computer Graphics Simulation of the Central Limit Theorem. *Mathematics and Computer Education, 21*(1), 48-55.

[12]. Groeneboom, P., Jong, P., Tischenko, D.B. & Zomeren, B. (1996). *Computer-assisted statistics education at Delft University of Technology, 5*(4), 386-399.

[13]. Kamtham, P. (1999). JAVA applets in education, irt.org.

[14]. Marasinghe, M.G., Meeker, W.Q., Cook, D. & Shin, T. (1996). Using Graphics and Simulation to Teach Statistical Concepts. *The American Statistician, 50*(4), 342-351.

[15]. Mittag, K. C., & Taylor, S. (1996). Using graphing calculator technology in educational statistics courses. Paper presented at *Annual Meeting of American Educational Research Association,* New York, April, 1996. (ERIC Document Reproduction Service No. ED 401 314)

[16]. Rosebery, A.S. & Rubin, A. (1990). Teaching Statistical Reasoning with Computers. *Teaching Statistics, 12*(2), 38-43

[17]. Segenchuk, S. (1997). *The Role of visualization in education.* Retrieved 6/12/2008:http://web.cs.wpi.edu/~matt/courses/cs563/talks/education/IEin

[18]. Velleman P.F. and Moore D.S. (1996) Multimedia for teaching statistics: promises and faults, *The American Statistician*, *50* (3), 217-225.

[19]. Windsor, N.J. (1998). Simulating a Sampling Problem Using a T1-83 Graphics Calculator. *Australian Senior Mathematics Journal, 12*(2), 32-37.

[20]. Zerbolio, D. J., Jr. (1989). A 'Bag of Tricks' for Teaching About Sampling Distributions. *Teaching of Psychology*, 16(4), 207-209.

ABOUT THE AUTHORS

Brandon K. Vaughn is an Assistant Professor in the University of Texas at Austin. His current research interests include multi-level differential item functioning (DIF), Bayesian estimation procedures, creative uses of non-parametric classification procedures, and effective strategies in the teaching of statistics. He has developed several technological tools for teaching statistics, including free R tutorial videos and applets for conceptual understanding.

Pei-Yu Wang is now a Doctoral Candidate in the Instructional Technology Program. Her current interests include interactive multimedia design, online system production, and open-source web development.