

GeoGebra in Teaching and Learning Introductory Statistics

Theodosia Prodromou
theodosia.prodromou@une.edu.au
University of New England
Australia

Abstract: *This article discusses the special opportunities for teaching statistics that technology offers teachers who aim to provide rich learning experiences for their students. These opportunities involve automation of many activities such as quickly organising data, computing measures, and generating graphs. By automating the tasks of computing statistics or generating data, technology facilitates students' ability to visualise abstract concepts, affording an opportunity to focus on conceptual understandings and data analysis. This article also examines how GeoGebra can be integrated into the curriculum and learning process of introductory statistics to engage college students in cycles of investigation including (a) managing data (b) developing students' knowledge for understanding specific statistical concepts, (c) performing data analysis and inference, and (d) exploring probability models. Recommendations are included for ways mathematics educators can engage college learners in developing their knowledge for exploring data concepts and doing statistics with GeoGebra.*

1. Introduction

We are all familiar with the increasing importance of technology within all realms of our existence. Technologies for doing and teaching statistics, such as graphing calculators, spreadsheets, and statistical packages such as SPSS, SAS, S and S-PLUS, and more recently R or Minitab, have become commonplace in many tertiary and some secondary contexts. This article discusses (1) the special opportunities for teaching statistics that technology offers teachers who aim to provide rich learning experiences for their students (2) how GeoGebra is designed for engaging college students in managing data, performing data analysis and inference, and exploring probability models is integrated into the curriculum and learning process of introductory statistics.

GeoGebra is of particular interest to educators with a limited budget because it is an open-source software package, available free on the internet and it can be an effective tool to teach geometry, calculus, algebra, and statistics.

2. What does technology offer teachers of statistics?

Pratt et al. [1] consider five affordances that technological advances offer to teachers of statistics as follows: (1) Using representations as dynamic tools for analysis, (2) expressing personal models, (3) expressing models, (4) storing and processing real data, and (5) sharing and communicating.

Using representations is the first affordance. Traditionally, graphs are used to report data, often through graphical displays and representations. Statistical software can provide graphical representations of data that are easily generated. Technology, therefore enables representations to be used as analytical tools at the end of the investigation. Pratt et al. [1] argue that “when graphs are used to try to make sense of data during analysis, the representations need to appeal to an intuitive sense of position, spread and outlying values. One example is the hat plot in TinkerPlots [2] designed to appeal to students' intuitive notion of modal clump [3]” (see [1], p. 98).

Pratt et al. [1] consider as a second important affordance for teaching statistics, the creation and use of models that capture the inherent structure in a situation. Teachers of statistics might therefore engage their students in an activity of using computational tools to express personal

models that better fit data arising from the random phenomenon in question. Computational tools empower the students' levels of expressiveness needed to create their own informal models for the data by looking for patterns in the data, a process often termed as expressive modelling [4].

Teachers could also use statistical software in which they build models that students explore by seeing sample data. According to Pratt et al. [1] exploring models is the third important affordance for teaching statistics. The process of engaging in exploratory modelling [4] and experimenting with statistical distributions [5] might provide opportunities for students to appreciate the use of key statistical concepts in particular contexts. When students are engaged in activities of expressive modelling, there is no guarantee that they are also engaged in constructing any key statistical concepts. Simulations have been used to help students draw connections between ambiguous statistical concepts such as epistemological perspectives for probability [6], unexplained and explained variation [7], or deterministic and stochastic causality [8].

Storing and processing real data is the fourth important affordance for teaching statistics [1]. Digital technology facilitates the storage and easy retrieval of large data sets. Use of large data sets enables the analysis of data drawn from real situations that are meaningful for students. Real data sets present issues that are usually not present in carefully prepared situations. These issues are difficult numbers, errors in data, and missing values. Such issues need to be confronted because they raise important points about the limitations and reliability of statistical inferences.

Sharing and communicating is the fifth fundamental affordance for teaching statistics with technology [1]. Several researchers argued that new technologies enhance the sharing and communication of information about statistics. For example, [9] has researched several types of wiki activities for the creation of resources that could be beneficial to teacher trainees of statistics and their future students. Amongst these activities, are collaborative writing, glossaries, discussion and reviews, statistical projects, self-reflective journals, and assessment. [10] investigated how a very large data set consisting 9000 email messages that contained 30 variables each was used to introduce teacher trainees to a dynamic document environment that included a live worksheet that in turn allowed students to interact with the data, adjusting inputs, altering computations and exploring a variety of different statistical analyses. A live worksheet with data allows users to reflect on what statisticians might do in order to perform a statistical investigation. Interactive environments reinforce reflection, which supports teachers in more effectively teaching statistics and students in drawing better connections between problems-solving activities in statistics and real life situations.

These five affordances that focus on using representations as dynamic tools for analysis, to express models, including personal models, to store and process real data, to share and communicate will be used to guide our analysis of GeoGebra. We might expect that technology affordances as described above [1] have changed the way teachers teach statistics and students learn statistics [1]. Technology in general and GeoGebra in particular, can help with amplifying ([12], [13]) users' abilities to solve problems or reorganise the way people think about problems in Statistics and search for their solutions.

We will reflect on a lecturer's use of GeoGebra [14] in teaching and learning Introductory Statistics.

3. GeoGebra

GeoGebra is a dynamic and interactive geometry, algebra, and calculus application, intended for teachers and students. Most parts of GeoGebra are free software and as of 2013 March, there are more than 120 GeoGebra Institutes (IGI) around the world. GeoGebra Institutes are non-profit

organizations at schools and universities who share free educational materials, support teachers and students in using GeoGebra, and work on projects related to GeoGebra.

GeoGebra was initially designed for teaching and learning geometry and algebra. Hence, it involves the following main features: dynamic geometry environment, built-in CAS, scripting, and a built-in spreadsheet with extended S. Geometrical constructions can be made using coordinates, to determine the position of points, vectors, segments, lines, polygons, conic sections, inequalities, implicit polynomials, and functions. Elements can be entered and changed dynamically on screen, or through the Input Bar. GeoGebra linked the learning Geometry, algebra, and calculus by using variables for numbers, vectors, and points to view the algebraic formula, find derivatives and integrals of functions (calculus), and it has a full complement of commands like Root or Extremum. Teachers and students can use GeoGebra to make conjectures and to understand how to prove geometric theorems because they conceptualise the mathematical formulae in relation to their geometrical constructions.

GeoGebra has very recently offered some tools that can be used for teaching statistics. We will reflect on how GeoGebra could be integrated into the curriculum and learning process of introductory statistics.

4. Teaching introductory statistics with GeoGebra

We will consider two distinct ways GeoGebra can be used when teaching introductory statistics courses:

- (1) Applets created with GeoGebra are integrated into lectures to demonstrate specific concepts.
- (2) Students use GeoGebra as a software tool to perform data analysis and inference and to explore probability models.

These applets are used during lectures and tutorials to demonstrate specific fundamental concepts that are commonly difficult to conceptualise in relation to a bivariate distribution. For example, the least Squares line demonstration introduces students to the standard approach to the approximate solution of sets of equations in which there are more equations than unknowns. The applet of least Squares line demonstration enables students to interact with the main characteristics of the “least squares” by dragging the scatterpoints in a graph window or the two points on the fit line, or overlaying the graph with statistical measures (e.g., residuals, squared residuals, means, a least square line and observing the augmentations of the graph. Users can change the data by dragging points in the graph, and the relevant statistical measures update automatically. Augmentations can be made to an existing representation to (a) alter representations, (b) overlay other characteristics other than statistical measure and additional information in the representation (e.g., a residual plot to show a graph of residuals) and visualise how the overall minimizes the sum of the squares of the errors made in the results of every single equation (Figure 4.1). This visualisation is not possible without technology that can scaffold students to reorganise their conceptions of bivariate distributions.

Another applet created with GeoGebra is the Student's t-distribution vs. Z demonstration. The applet demonstrates the Student's t-distribution that has the following probability density

function,
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$
 where ν is the number of degrees of freedom and Γ is the gamma distribution. The applet provides students the ability to adjust the degrees of freedom (by moving a slider) and observe the altered graphical representations that show how the t-distribution

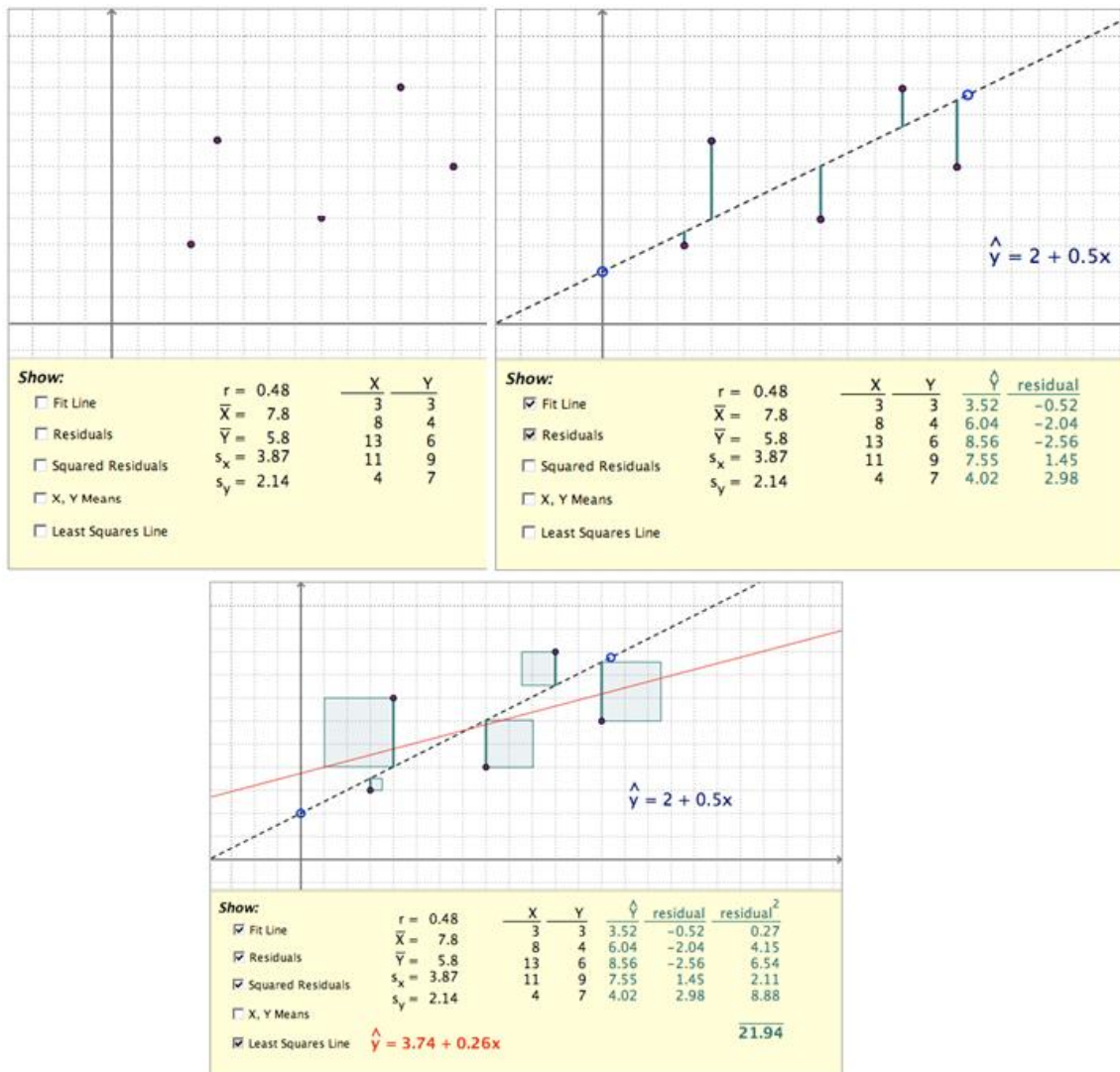


Figure 4.1 The applet of least Squared line demonstration

becomes closer to the normal distribution (with mean 0 and variance 1) as the number of degrees of freedom increases.

The following representations (Figure 4.2) show the density of the t-distribution (red line) for increasing values of v . The normal distribution is shown as a blue line for comparison. Figure 4.2 also illustrates how to create dynamic links among the t-distributions and the Normal Distribution. A dynamic link, in contrast to a static link, “occurs when a purposeful action is done to one representation that causes a reactive and coordinated action in another representation” [15]. A static link within a dynamic software, occurs when there is an indication of coordinating complementary information in two or more representations [16] that does not involve any direct technological interaction with a representation.

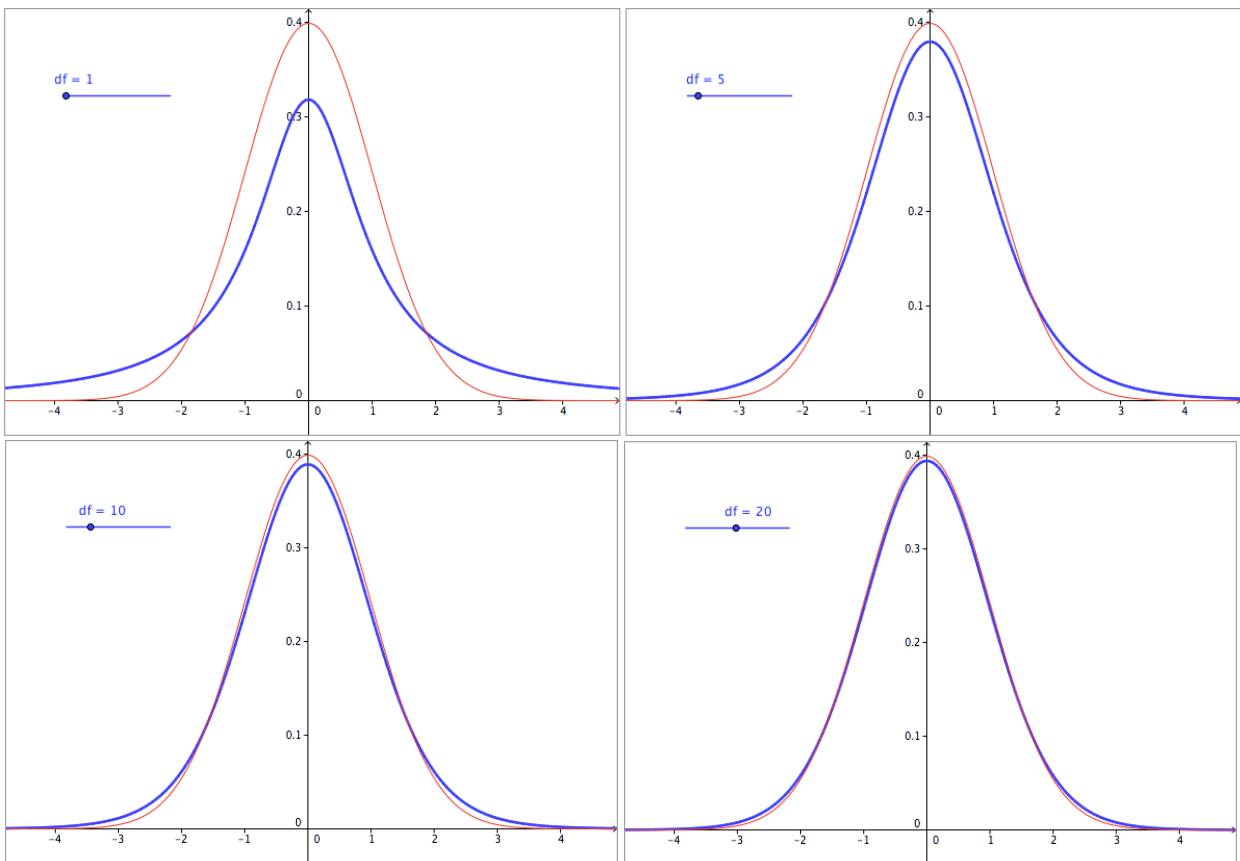


Figure 4.2 Density of the t-distribution (red) for 1, 5, 10 and 20 degrees of freedom compared to the standard normal distribution (blue).

Ultimately, GeoGebra applets like those shown above provide affordances of dynamic statistical software such as abilities to relate the two representations, connecting them dynamically, and obtaining simultaneously graphical augmentations of the density of the t-distribution (red) for 1, 5, 10 and 20 degrees of freedom. Taking advantage of these affordances users could actively be engaged in statistical problem solving activities that may lead to further investigations or additional insights.

In this study we are also interested in how teachers can take advantages of the features in GeoGebra that allow some of the affordances described by [1] in their teaching of introductory statistics to engage college students in (a) managing data, (b) performing data analysis and inference, and (c) exploring probability models. A description of (a) to (c) follows.

5. Using GeoGebra as a statistics software package

GeoGebra can be used as a statistics software tool to help students perform data analysis and inference and explore probability models. Students can work with a specially configured version of GeoGebra designed for a statistics-learning environment (Figure 5.1). This specially configured version of GeoGebra is only accessible by the students of the lecturer ¹who has designed this version of GeoGebra.

¹ George Sturr has designed this version of GeoGebra.

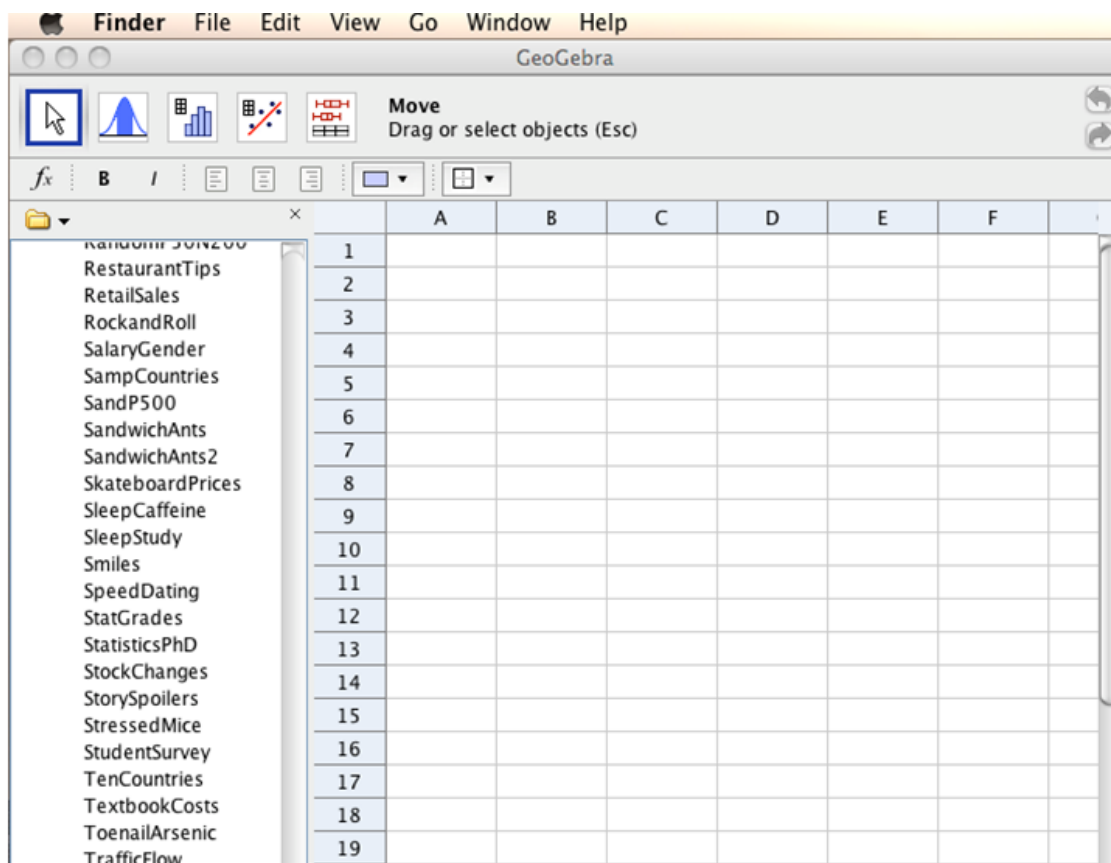


Figure 5.1 The version of GeoGebra

In this example, the students do not use the input bar or graphics window. All work is done from the spreadsheet and data analysis dialogs. This action of using data sets of data points from a spreadsheet and data analysis dialogs is the primary building tool of GeoGebra when creating the representation of data analysis tools to establish a link between the data analysis graphs or summary tables to the data that is stored in the spreadsheet table. The users of GeoGebra view a spreadsheet table of data points and a graphical representation separately. This capability of storing and processing of (real) data reflects the fourth affordance described by [1]. Other affordances described by [1] are exhibited such as the use of graphical representations of analysing data (first affordance) while students express models (third affordance) that describe the data, including personal models (fourth affordance).

A student may link the information from the data points in the spreadsheet table with the information in each plot. This way of coordinating information is done by establishing a static link between the data points of one variable and the information in each plot or summary table. The structure of the GeoGebra environment does not allow representation of data to be linked as they are created as other dynamic environments do (e.g., Fathom or TinkerPlots). In Fathom[17] or TinkerPlots[2], when creating a graphical representation or a summary table, one can drag an attribute name (variable) from the data card or data table onto a plot window or an axis in a graph window or a row or column in a summary table.

Students and lecturer meet in a computer lab and students use GeoGebra regularly during class meetings. Homework assignments that require data analysis, inference, or probability calculations are expected to be done with GeoGebra.

5.1. Managing Data

Data sets are re-entered and manipulated as columns in GeoGebra's spreadsheet. Data can be imported into the spreadsheet with cut and paste techniques or by opening data files from a local drive or external web site.

Accessing data files is greatly simplified with GeoGebra's data browser, which displays a catalog of locally stored data files or a catalogue of files stored on a web page. With one click on a file title the spreadsheet is cleared and the data file imported. In this example, the data browser is set by default to list the data files for the course textbook examples and homework problems.

This optional sidebar panel provide students with the capability to quickly open and explore data files spontaneously in response to student homework questions. Storing and processing data as well as analysing and displaying data have the features of the four affordances as described by [1]. GeoGebra also facilitates the use of large data sets of interest to students that are collected and accessed through internet. Knowing how to collect real data from the internet and how to transform it into usable data in a particular piece of software (e.g., GeoGebra) is a useful skill. The use of discussion forums, wikis, interactive whiteboards, and self-assessment software can promote collaboration and students' involvement, supporting a focus on collaborative tools for data collection, analysis, and data visualisation. Collaborative tools to collect data can promote a community approach to generating and analysing data that can foster sharing of information and collaborative investigations of data and discussion to consider the statistical problems. This reinforcement of sharing and communicating constitutes the fifth affordance described by [1].

5.2. Creating Graphs and Performing Inference

GeoGebra provides students with opportunities to create graphs as tools for analysis or perform inference with GeoGebra. To perform inference with GeoGebra, a student selects data from spreadsheet columns and presses a toolbar button to launch a data analysis dialog window. The dialog opens with a simple graphical representation of the data and offers options to display statistical measures and sub-dialogs for calculating confidence intervals and hypothesis tests (Figure 5.2).

There are three data analysis options listed as follows: One Variable Analysis, Regression Analysis, and Multi-Variable Analysis.

The One Variable Analysis button is used for single variable data, both quantitative and categorical. It generates bar charts, histograms, boxplots, stem and leaf plots, dot plots, or normal quantile plots. The Regression Analysis button is used for quantitative bivariate data and generates scatterplots or residual plots with a variety of regression models. The Multi-Variable Analysis button generates stacked boxplots and one-way ANOVA tables.

For some dialogs there is a data-selection panel for setting data values. This option often affords opportunities to consider a view of a distribution of a group of data values and juxtaposed this view to an aggregate view of a distribution. For example, the students might use graphical representations to consider the impact of particular data values in a distribution of aggregate values or to examine the impact on a measure of center if an outlier is removed from a data set. GeoGebra also allows students to focus on an interesting subset of data by allowing them to segment a distribution's range into several parts, and group data visually within a range in order to examine special features in subregions of a distribution.

5.3. Probability and Statistical Inference Calculator

GeoGebra provides students with its Probability Calculator (Figures 5.2 and 5.3). Students can use this feature to explore data, calculate probability models, and make claims and decisions

about basic statistical inferences from data. The software tools in this dialog are equivalent to or surpass those found in a graphing calculator.

The Probability Calculator within GeoGebra has two tabbed panels, Probability and Statistics. The Probability tab panel allows students to view a variety of common probability models. In particular, it displays a graph for a selected model and has a panel with a simple, button-driven interface for computing interval probabilities (Figure 5.2). The graph gives visual feedback

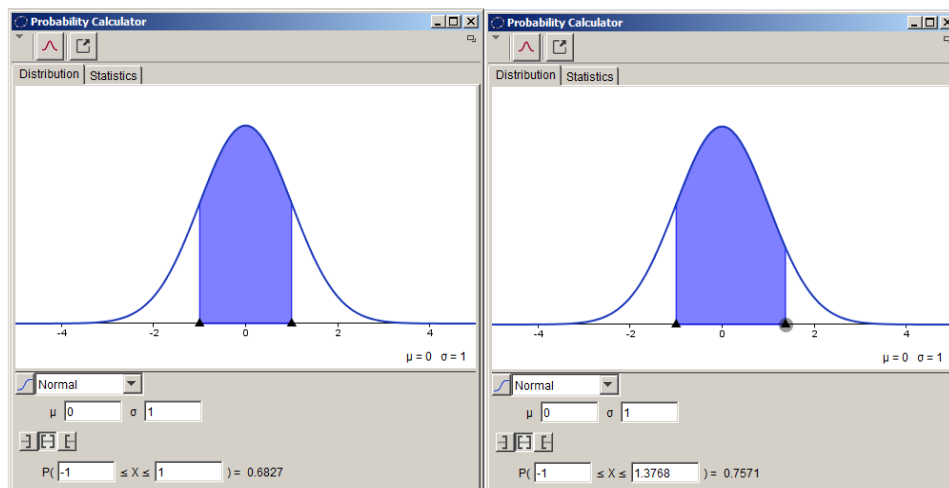


Figure 5.2 Probability Tabbed panel of Probability Calculator

for computed probability intervals and allows for dynamic changes by dragging interval endpoints. This dynamic dragging of interval endpoints allows students to establish a dynamic link coordinating a single graphical representation of a distribution with statistics like mean, variance, and standard error, as well as tools for different hypothesis tests. Students are able to use dynamic linking to look at basic probability models and make inferences from data by performing hypothesis testing and observing every time how the values of mean, variance, standard error are modified (Figure 5.3). Students often compare the impact of modifying a certain model to a statistic and using such comparisons they can make statements about relationships.

Students establish dynamic links when coordinating information about graphs of a selected model and computed probability intervals, which allows for dynamic changes by dragging interval endpoints. Students usually made static links toward the end of making concluding claims and decisions about basic statistical inferences from data, as they reflected on their interactions with the tools of the probability and statistical inference calculator.

The Statistics tab panel is a dialog for computing common confidence intervals and hypothesis tests from summary statistics. These software calculator tools replace table lookups and hand-held calculators. The majority of the college students already own a hand calculator that can do statistical work.

6. Discussion

GeoGebra provides the five affordances that Pratt et al. [1] suggest that technological advances offer to users. In GeoGebra, representations of data are used as dynamic tools for analysis at the end of each investigation when solving statistical tasks. This feature of GeoGebra matches

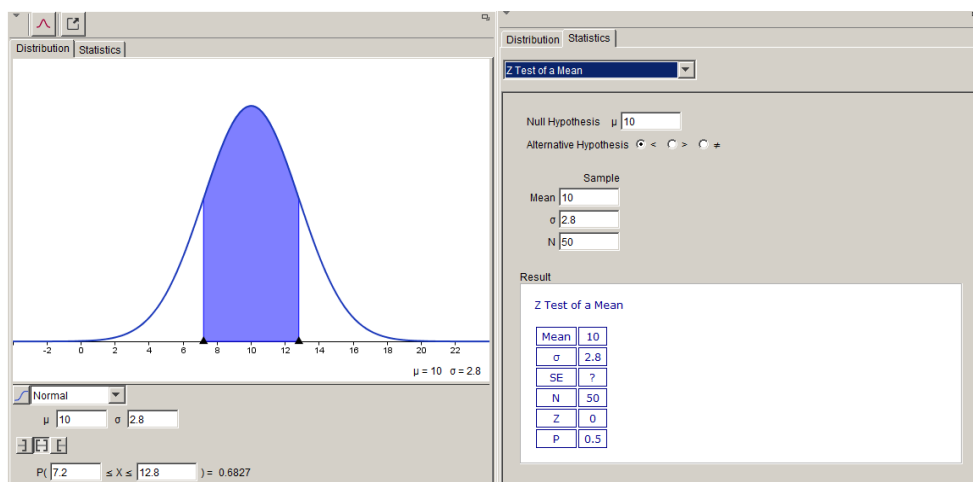


Figure 5.3 Statistics Tabbed panel of Probability Calculator

Pratt's first affordance. The representations that are available in GeoGebra however, are lacking three important structures of dynamic environments. First, the graphical representations are created without establishing a connection between the summary table and the graph to the data that is stored in the data table. Hence, the graphical representations of data (e.g., tables, graphs) are not associated with each other while they are constructed because when one creates a graphical representation or a summary table, he/she cannot drag a name of a variable from the data card or data table to drop it onto an axis in a graph window or a row or column in a summary table.

Second, GeoGebra does not provide yet tools that could have an appeal to students' intuitive sense of some statistical concepts. It would be better for middle school students and younger users of the software to have access to individual data values so they could interact directly with them. For instance, dragging a data point within a dot plot could contribute to instantly separating data values into groups for the same variables and generating common graphical displays of those variables. Students could also actively construct different representations by dragging the name of the representation on to a selected representation. Third, in GeoGebra when highlighting a data point in one representation, it does not create a highlighted display of that data point within all other representations. GeoGebra does not yet include built-in connections among all representations created from a data set.

GeoGebra affords users with the fourth affordance pointed out by [1], in that it facilitates real data storage and processing of real data drawn from situations that are meaningful for the students. It can be used by students to process data in a faster way, calculate sophisticated statistics, analyse data, and draw appropriate inferences. The way that GeoGebra supports users to explore probability models places emphasis on the third affordance of modelling as well as on data analysis as discussed above. This feature of GeoGebra also matches Pratt's fifth affordance of sharing and communicating statistical results through discussion forums, wikis, interactive whiteboards, and self-assessment software to promote collaboration and students' involvement.

The Probability tab panel in the probability calculator could help users to view several graphical representations of different probability models, offering students a range of affordances that could scaffold their learning of statistics, ensuring that students could link statistical thinking to probabilistic thinking.

GeoGebra indisputably offers a range of affordances as described by [1], and technological advances can help provide those affordances, with GeoGebra as an example. The development of statistical tools in GeoGebra, will of course continue, and no doubt so will the careful analysis of

how its design and implementation have an impact on mathematical teaching and learning. Such developments of mathematical software need to be examined and evaluated to help mathematicians and software designers to better understand the consequent impact that results from learners' interaction with the software; and the way that teachers integrate the special characteristics of GeoGebra into teaching to support students' learning of statistical ideas.

References

- [1] Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 97-107). New York: Springer Science+Business Media B.V. 2011.
- [2] TinkerPlots: *Dynamic data exploration* (Version 2.0) [Computer software]. Emeryville: CA: Key Curriculum Press.
- [3] Konold, C., & Pollatsek, A. (2002), Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33, 259-289.
- [4] Doer, H., & Pratt, D. (2008). The learning of mathematics and mathematical modelling. In M. K. Heid & G. W. Blume (Eds.), *Research on technology in the teaching and learning of mathematics: Syntheses and perspectives. Mathematics learning, teaching and policy* (Vol. 1, pp. 259-285). Charlotte, NC: Information Age.
- [5] Garfield, J., & Ben-Zvi, D. (2008). Preparing school teachers to develop students' statistical reasoning. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (2008).
- [6] Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools. *International Journal of Computers for Mathematics Learning*, 12(1), 23-55.
- [7] Engel, J., Sedlmeier, P., & Worn, C. (2008). Modelling scatter plot data and the signal-noise metaphor: Towards statistical literacy for pre-service teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (2008).
- [8] Prodromou, T., & Pratt, D. (2006). The role of causality in the coordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal*, 5(2), 69-88. Online: www.stat.auckland.ac.nz/serj/
- [9] Ben-Zvi, D. (2007). Using wiki to promote collaborative learning in statistics education. *Technology Innovations in Statistics Education*, 1(4). Online: repositories.cdlib.org/uclastat/cts/tise/
- [10] Nolan, D., & Temple Lang, D. (2007). Dynamic, interactive documents for teaching statistical practice. *International Statistical Review*, 75(3), 295-321.
- [11] Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, 1(1). Retrieved from <http://escholarship.org/uc/item/8sd2t4rr>
- [12] Pea, R. D. (1987). Cognitive technologies for mathematics education. In A. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 89-122). Hillsdale, NJ: Erlbaum.
- [13] Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning*, 2(1&2), 127-155.
- [14] Hohenwarter, 2001 GeoGebra. Available from www.geogebra.com
- [15] Lee, H. S., Kersaint, G., Harper, S., Driskell, S. O., & Leatharm, K. R. (2012). Teachers' statistical problem solving with dynamic technology: Research results across multiple institutions. *Contemporary Issues in Technology and Teacher Education*, 12(3). Retrieved from <http://www.citejournal.org/vol11/iss3/mathematics/article1.cfm>
- [16] Ainsworth, S. E. (1999). A functional taxonomy of multiple representations. *Computers & Education*, 33, 131-152.
- [17] Finzer, W. (2002). *Fathom dynamic data software* (version 2.1) [Computer software]. Emeryville, CA: Key Curriculum Press.