The STAT-ATTIC Website: Links to Statistics Applets for Introductory Courses

Concetta A. DePaolo
Indiana State University

Key Words: Statistics Education; Computers; Technology.

Abstract

Statistics instructors are increasingly using stand-alone applets to illustrate statistics concepts to students. There are hundreds of applets available on the web, but locating a quality applet to serve a particular purpose can be time consuming. This paper presents the STAT-ATTIC (STATistics Applets for Teaching Topics in Introductory Courses) website, located at http://sapphire.indstate.edu/~stat-attic/index.php, as a tool for locating statistics applets for a particular pedagogical purpose. The site has links, with descriptions, to approximately 600 publicly available applets on topics commonly taught in introductory courses, including graphical displays, descriptive statistics, probability concepts, random variables, sampling, confidence intervals, hypothesis tests, regression, ANOVA and chi-square tests. Within each broad topic, applets are organized by intended purpose; for example, applets intended to illustrate properties or behavior are shown separately from those that calculate statistics or perform tests. Advantages of the database over other statistics resources are presented, along with the process that was used to identify and select applets for the site. Lastly, the structure and functionality of the database, including the ability to search and rate or comment on applets, are also discussed.

1. Introduction

I have taught introductory undergraduate business statistics courses at a mid-sized Midwestern public university for more than ten years. The two-course sequence consists of a fairly typical first course in statistics that covers graphical displays, descriptive statistics, basic probability concepts, random variables, sampling, and confidence intervals and hypothesis tests for one and

two populations. The second course is somewhat unusual, covering not only ANOVA, chi-square tests and simple and multiple linear regression, but also forecasting and time series, decision analysis, and an introduction to calculus.

Over the years, my approach to the technology component of the courses has changed drastically. My sections of 30-45 students originally were taught in a regular classroom, with four to five class periods held in computer labs to give students hands-on experience with using Excel and Minitab to do statistical calculations and analysis. However, in 2008, our university began requiring that all students have laptops. I recognized that my courses could (and should) undergo a drastic pedagogical shift. I could move away from lecture format with practice, discussions, computer demonstrations by the instructor and occasional trips to the lab, to a hands-on, active learning environment in which students could immediately work with the data, perform calculations, and visualize concepts right at their own computer screens. This was an exciting, but daunting, proposition.

I set out to transform my courses by incorporating some principles from the American Statistical Association's GAISE (Guidelines for Assessment and Instruction in Statistics Education) College Report. This report gave several recommendations for reforming statistics instruction, one of which was to "use technology for developing conceptual understanding and analyzing data." Specifically, the report suggested several ways to effectively use technology, including the use of large datasets, automation of calculations, generation of statistical graphics, performing simulations to illustrate concepts, and performing "what-if" analyses (GAISE, 2005). I began searching for ways to incorporate these ideas into my courses. Since one of our college's learning goals is proficiency with technology, with specific emphasis for our business students on Excel and Access, I envisioned using Excel for some of these activities. On the other hand, I also wanted to find specialized computer applications to do simulations and illustrate statistical concepts. I began searching for these applications, mostly Java applets, in 2009.

I began my search by reviewing articles in statistics teaching journals. I also searched through MERLOT (http://www.merlot.org/) and Causeweb (http://www.causeweb.org/), which both have wonderful resources for statistics instructors but required me to wade through a lot of other instructional materials besides applets that I did not necessarily need. I did find several nice applets and applet collections, for example:

- The Rossman/Chance collection (http://www.rossmanchance.com/applets/index.html),
- SOCR (http://www.socr.ucla.edu/SOCR.html),
- Statistical Java (http://www.causeweb.org/repository/statjava/),
- Rice Virtual Laboratories Statistics (http://onlinestatbook.com/stat_sim/index.html),
- Vassar Stats (http://faculty.vassar.edu/lowry/VassarStats.html), and
- A Collection of JavaScript E-labs Learning Objects (http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/scientificCal.htm).

The applets in these collections did not always do exactly what I wanted or did not have applets to address all of my needs. When I did find applets, I sometimes noticed that there were several different similar applets that had slightly different functionality, and I was unsure exactly which I would use when the time came. I also spent some time googling various phrases (statistics

applets, regression applets) and got several hundreds of thousands of hits each time, but many were for commercial products. I did find several sites with links to multiple applets, some even organized by topic, but unfortunately, there were often limited, if any, descriptions of the functionality, and in many cases, a good number of broken links.

I realized fairly quickly that finding what I was looking for was far more time-consuming than I anticipated. I was not going to be able to prepare the course "on the fly" by finding an applet (or several) that I needed for a particular purpose as I completed my preparation for the next class period. I decided to start cataloging the applets that could be useful in my courses or similar courses that I liked and were easy for students to use. I began organizing them by topic in an Excel database.

I estimate that I evaluated between 1000-1500 applets. When I concluded my search, I ended up with a teaching arsenal of over 600 applets, organized into these 13 topics:

1. Graphical Displays of Data,
2. Descriptive Statistics,
3. Probability Concepts,
4. Random Variables & Probability Distributions,
5. Sampling & Sampling Distributions,
6. Confidence Intervals,
7. Hypothesis Tests,
8. ANOVA,
9. Chi-square Tests,
10. Regression,
11. Time Series & Forecasting,
12. Decision Analysis, and
13. Quality Control Charts.

Within each of these 13 broad topics, I also included in the database more narrow categories that would align with how the applets would be used. For example, a regression applet could calculate statistics or demonstrate the least squares concept. An applet under hypothesis testing could either illustrate basic concepts, simulate or illustrate p-values, or conduct tests for one or two populations. There are a total of 55 subtopics within the 13 main topics of the database. (For a complete list of subtopics, refer to the appendix.) Each applet in the database is catalogued under the appropriate topic and subtopic with its source, name, url, description, and appropriate search terms. Some applets that perform multiple functions may be listed under more than one topic; for example, an applet may both calculate descriptive statistics and confidence intervals and so is cross-listed under both topics.

In early 2010, our university's instructional design programmers developed a web interface for the database I had created, and STAT-ATTIC (STATistics Applets for Teaching Topics in Introductory Courses) came to life. In the remainder of this paper I describe why and exactly how I think STAT-ATTIC can be of use to other instructors. I then discuss how applets were selected for inclusion in the database and conclude with a description of the organization and functionality of the STAT-ATTIC website.

2. Why and How to Use STAT-ATTIC

I admit that STAT-ATTIC may not be extremely useful to some instructors. If you are looking for a well-developed instructional activity or materials, then the applets in STAT-ATTIC are probably not for you. Instead, a search through Causeweb (http://www.causeweb.org/) and MERLOT (http://www.merlot.org/) would be more beneficial.

I use applets to illustrate concepts that I have taught for years with my own examples, worksheets and student activities. I apply the technology to calculate or illustrate something that I previously would have done myself with Excel or by hand. For instructors who are looking for such "plug-ins" to supplement their own instructional materials and know exactly what they want or need, then STAT-ATTIC is a good resource to help you find it quickly.

One advantage of STAT-ATTIC over CauseWeb and MERLOT is that applets in STAT-ATTIC are organized not only by broad topic (e.g. regression or ANOVA), but also by subtopics that describe exactly how the applet is to be used. For example, if you are looking for a regression applet to illustrate the least squares concept, you can view in STAT-ATTIC only those applets that perform that function, as opposed to wading through several regression applets that may only calculate regression statistics or correlations.

Furthermore, STAT-ATTIC's descriptions allow easier side-by-side comparison of similar applets. For example, the database currently contains information on nine applets that conduct single sample hypothesis tests. Users are able to view which parameters may be tested, how many observations the applet can handle, whether the data is input one by one or can be pasted in, and whether it does both one- and two- tailed tests. By having this information readily available, the user avoids having to click on each applet before realizing that it does not suit his or her needs. In addition, even if several are suitable, since the links are all gathered in one place, it is easy for the user to compare them and quickly select one.

Having links to several similar applets all together is also an advantage of STAT-ATTIC over some of the very fine applet collections that are available. While many of these collections have useful and high quality applets, not every instructor will find everything he or she is looking for in a single collection. Perhaps one collection has a Central Limit Theorem applet that is particularly useful, but it does not have a confidence interval applet that suits one's particular needs. By drawing on applets from many collections and sources, the user has a single resource for locating exactly what he or she requires for many different topics.

Lastly, by drawing on many resources, including clearinghouses, collections, and independently developed applets, STAT-ATTIC casts a wider net than some other resources. A search through CauseWeb for the word "applet" results in 261 hits, and the same search within the "Statistics and Probability" learning materials in MERLOT returns 70 hits, compared to almost 600 STAT-ATTIC applets. As another example, a more refined search for "regression applet" yields 29 and 7 hits in CauseWeb and MERLOT, respectively. This compares to STAT-ATTIC's 70 regression applets, which furthermore can be searched or browsed by subtopic or intended use.

Obviously, one limitation of STAT-ATTIC is that the applets included have not yet been evaluated and rated by users other than myself. The database does include a rating and feedback function, so this information can be collected and eventually reported back to users, but as of now, there are no external ratings. However, none of the 261 applets that are found in a search for "applet" in CauseWeb currently have any ratings available. Of the 70 statistics and probability applets in MERLOT, only about half are rated at all, and the vast majority of those have only been reviewed by one peer. Thus, it appears as though there is a need for users to share feedback on statistics applets found in many sources to help others identify quality products.

3. The Applet Selection Process

I began the search for applets in two ways. First, I reviewed statistics teaching journals, including Teaching Statistics, Statistics Education Research Journal, Technology Innovations in Statistics Education and The Journal of Statistics Education. JSE has an Archive of Interactive Computing (http://www.amstat.org/publications/jse/jse_computing_archive.htm), which includes several nice applets. The papers and applets I found often referenced other sites and applets, which I then reviewed. Second, I scoured through clearinghouses of statistics teaching materials, including MERLOT (http://www.merlot.org/) and Causeweb (http://www.causeweb.org/) through which I located some quality applet collections, including SOCR (http://www.socr.ucla.edu/SOCR.html), Statistical Java (http://www.causeweb.org/repository/statjava/), and the Rossman/Chance collection (http://www.rossmanchance.com/applets/index.html) in addition to others previously discussed. These sites sometimes led to other sites and collections. When I had exhausted these avenues, I finished with a Google search, which sometimes identified appealing applets I had not found elsewhere.

I used several criteria when selecting applets for the database.

Stand-alone. The applets included had to be stand-alone applets. There are some quality online textbooks that make extensive use of applets (e.g. CAST http://cast.massey.ac.nz/collection_public.html and Hyperstat Online http://davidmlane.com/hyperstat/index.html); however, because the applets are embedded in the textbook itself, it is often not easy to locate them within the text and they often appear with other material that I do not emphasize. On the other hand, Seeing Statistics (http://www.seeingstatistics.com/) does embed applets in the text, but they can also be accessed individually and so several of these applets are included in the database.

No downloads. Generally, I did not include applets or applications that involved downloading software because I felt that it was not reasonable to ask students to download and install software during a class period, especially if I could find another easily accessible applet that did essentially the same thing. One notable exception involves several Wolfram Demonstrations (http://demonstrations.wolfram.com/topic.html?topic=Statistics), which give very nice visuals to explain statistical concepts. I regularly use these in my courses as a demonstration, rather than having students download and play with the demos themselves.

Free. There are many commercial products that will do statistical calculations, but if there was any cost to using the application, it was not included in the database.

After eliminating products or applications that did not meet these general criteria, I turned to the issue of desirability. Following some general guidelines as described in Smith (2002), I used what I call the four 'A's for applet selection:

1. Applets had to be **appropriate** for introductory undergraduate courses. They had to address topics and concepts commonly taught in these courses at introductory levels. They also had to have functionality, for example, interactivity, that would allow for demonstration or exploration of concepts or calculation of statistics.

2. Applets had to be **accessible** to undergraduate non-majors. They had to have an intuitive interface, explanation of purpose, and brief but helpful documentation on usage. Applets that had an excess of options, inputs and functions were rejected in favor of simpler ones that undergraduate students would find easier to understand.

3. The applets' **attributes** had to be of high quality. They had to have reasonable loading and running speed and be free of errors and misleading output. Applets chosen were clear in their purpose and successfully performed their intended function or demonstrated their intended concept.

4. Applets had to be **appealing** visually. They had to have effective (though not necessarily fancy) graphics and produce output that is easy to understand and interpret.

Of the hundreds of applets I examined, many did not fit these criteria. In particular, there were several that performed identical functions to others, though in a way that was not as pretty, elegant, or efficient. These applets were not included in the database. However, there are several instances in which several equivalent applets are presented because none was inferior to the others and I considered the choice of using one or the other a matter of preference.

One final consideration is the reliability and accessibility of the applets. There were cases when I originally included an applet for the database but later found the link to be broken. These links have been removed, but of course, there are undoubtedly cases like this in the current incarnation of the database that I have not yet detected. Removal of broken links will be an ongoing process for the database. Hopefully the inclusion of multiple applets that perform similar functions will alleviate any issues for the user that broken links might cause, as an alternative will likely be found within the database.

4. The STAT-ATTIC Website

The STAT-ATTIC website (see Figure 1), located at http://sapphire.indstate.edu/~stat-attic/index.php, is a user-friendly tool for instructors to locate applets for specific purposes.

Figure 1

The site allows users to locate applets in two ways: either by browsing by topic or by searching for pre-set keywords using the "Search" function. As you type a keyword in the search box, a menu shows you all keyword choices that match your entry. For example, if you type "Regression" there are actually more than 10 topics to choose from (see Figure 2).
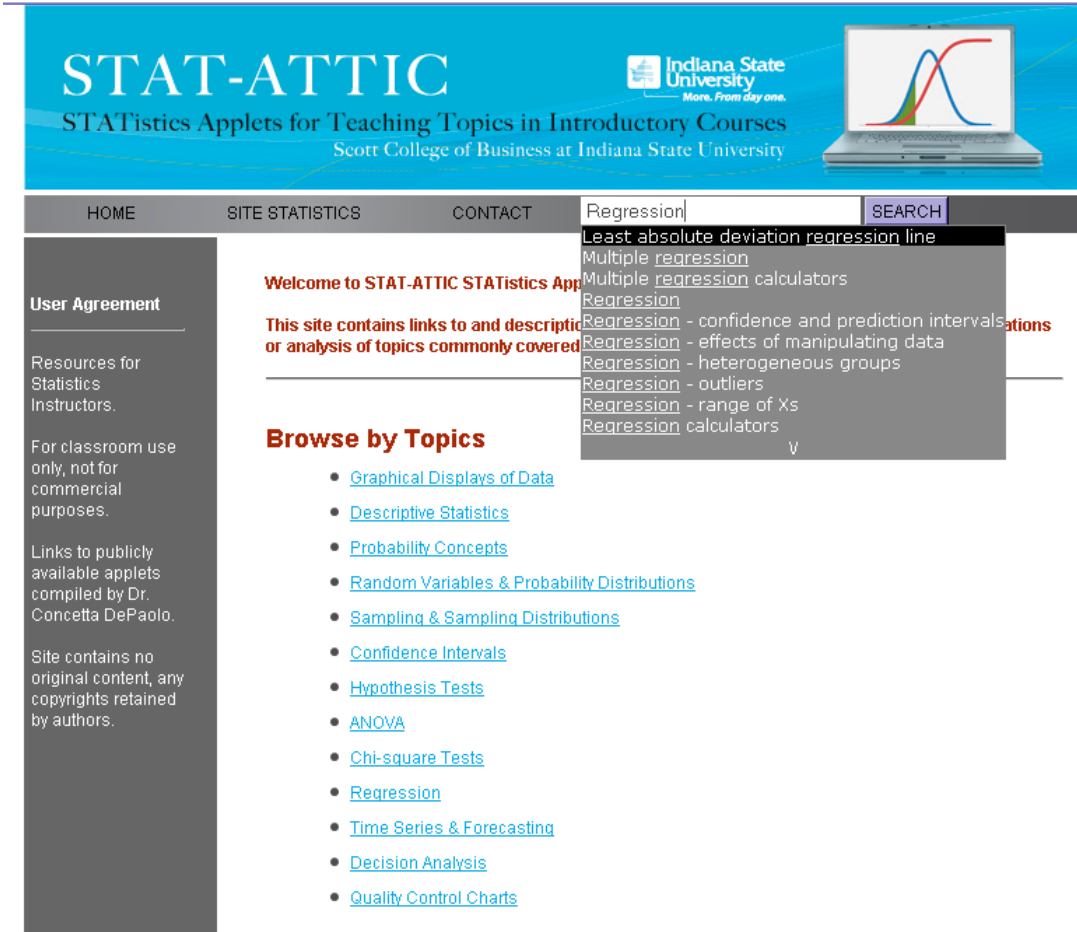
Figure 2

Once the relevant applets are located by topic or by keyword, they are shown in a list, sometimes on multiple pages, with the name, source, and a brief description of the functionality. For example, Figure 3 illustrates the first page of results if one selects "Confidence Intervals" and then the subtopic "Illustrate basic concepts and interpretation of confidence intervals." The "next" link at the top right indicates that the list is continued on subsequent pages.

Figure 3

If a user finds an applet of interest, he or she can click on a "more" link to bring up the item detail, including the url, keywords, cross listings (if applet is listed under more than one category) and number of times the applet has been visited. For example, Figure 4 shows the details that users see for the Normal Probability Calculator applet from the Rossman/Chance collection (found under "Random Variables & Probability Distributions" and the subtopic "Normal demos -- simulate or illustrate properties and behavior.") In addition, this detail page contains a comment box and rating form so that after a user visits an applet, he or she can comment on what they did and did not like about the applet. These comments are automatically emailed to me so that I can compile information on user likes and dislikes. I would encourage any users of the site to take the time to rate and/or comment on the applets. Although the capability does not currently exist to share these ratings and comments with users, eventually I would like to be able to include this functionality in the database as a way to help users identify desirable applets.

Figure 4

The site also contains a tab called "Site Statistics" that counts the total number of visits to the site and shows the 10 most visited applets in the collection. Lastly, contact information is included for users to provide general feedback, suggest links or applets, or report broken links.

Appendix: Complete List of STAT-ATTIC Topics & Subtopics

Graphical Displays of Data
1. Creating graphical displays
2. Evaluate, analyze and compare graphs

Descriptive Statistics
1. Calculate descriptive statistics
2. Evaluate, analyze and compare statistics

Probability Concepts
1. Construct Venn diagrams
2. Illustrate Independence, conditional probability
3. Illustrate or compute probabilities for Coins and/or Dice
4. Illustrate or compute probabilities for other models or experiments
5. Compute probabilities using contingency tables

Random Variables & Probability Distributions
1. Binomial demos – simulate or illustrate properties and behavior
2. Binomial calculators
3. Normal approximation to Binomial
4. Normal demos – simulate or illustrate properties and behavior
5. Normal calculators
6. T, F and Chi-square distributions
7. Other random variables and probability distributions
8. Probability distribution calculators for multiple distributions

Sampling & Sampling Distributions
1. Generate random numbers
2. Illustrate basic concepts of sampling and estimation
3. Simulate sampling distributions
4. Illustrate the CLT or size of standard error as sample size increases
5. Calculate standard error or probabilities for sampling distributions

Confidence Intervals
1. Illustrate basic concepts and interpretation of confidence intervals
2. Calculate single population confidence intervals or margins of error
3. Calculate two population confidence intervals
4. Calculate sample size necessary to achieve specified margin of error

Hypothesis Tests
1. Illustrate introductory concepts or steps of testing a hypothesis
2. Simulate or illustrate p-values, detection of differences, Type I and Type II errors, and/or power
3. Conduct hypothesis tests for a single population with user data
4. Conduct hypothesis tests for two populations with user data
5. Compute power of a test or determine sample size

ANOVA
1. Visually illustrate or simulate data to demonstrate ANOVA concepts
2. Calculate ANOVA table and statistics for user-entered data
3. Conduct post-hoc tests or test ANOVA assumptions

    4. For multiple comparisons, illustrate or calculate true error rates and effects of adjustments
    5. Two-way ANOVA: simulate or illustrate interactions and concepts, or calculate tables

Chi-Square Tests
    1. Simulate data to illustrate chi-square GOF or independence concepts, tests and results
    2. Conduct a GOF test with user-entered data
    3. Conduct a test of independence with user-entered data
    4. Conduct special analyses for 2x2 tables, e.g. Yates' correction, Fisher's exact test

Correlation & Regression
    1. Illustrate basic concepts of correlation
    2. Calculate sample correlations or test/inference on correlation based on user-entered data
    3. Illustrate basic concepts of the least squares regression line
    4. Illustrate how correlation and/or regression statistics are affected by points, outliers, range of Xs
    5. Simulate data with specified correlation coefficients, slopes, standard deviations, standard errors
    6. Calculate regression statistics or test the population slope based on user-entered data
    7. Calculate multiple regression statistics or illustrate multiple regression surfaces

Time Series & Forecasting
    1. Create a time series plot, calculate statistics, or test for stationarity, autocorrelation, or randomness of fluctuations
    2. Calculate or illustrate moving averages
    3. Calculate exponential smoothing forecasts or illustrate concepts
    4. Calculate trend and/or seasonal indices or test for their presence in a series
    5. Evaluate forecast accuracy

Decision Analysis
    1. Illustrate basic decision analysis concepts and decision strategies
    2. Decision analysis calculators for user-entered data

Quality Control Charts

Acknowledgements

References

GAISE (2005). Guidelines for assessment and instruction in statistics education (GAISE) college report. Retrieved May 14, 2010, from The American Statistical Association (ASA): http://www.amstat.org/education/gaise/GAISECollege.htm.

Smith, B. E. (2002). Statistical Applets: The Good, the Bad, and the Ugly. Retrieved April 15, 2010, from Borough of Manhattan Community College:
http://www.bmcc.cuny.edu/mathsci/2002/presentations/briansmith.html

Concetta A. DePaolo, Ph.D.
Associate Professor
College of Business
800 Sycamore Street
Indiana State University
Terre Haute, IN 47809
Email: Concetta.DePaolo@indstate.edu
Phone: (812) 237-2283

Volume 18 (2010) | Archive | Index | Data Archive | Resources | Editorial Board | Guidelines for Authors | Guidelines for Data Contributors | Guidelines for Readers/Data Users | Home Page | Contact JSE | ASA Publications

13