Simulation and Bootstrapping for Teaching Statistics

Tim C. Hesterberg
MathSoft, Inc., 1700 Westlake Ave. N., Suite 500
Seattle, WA 98109-3044, U.S.A.
TimH@statsci.com, http://www.statsci.com/Hesterberg

Abstract
Some key ideas in statistics and probability are hard for students, including sampling distributions. Computer simulation lets students gain experience with and intuition for these concepts. Bootstrapping can reinforce that learning, and provide a way for students (and future practitioners!) to estimate sampling distributions when they have data but do not know the underlying distribution. Bootstrapping also frees us from the requirement to teach inference only for statistics for which simple formulas are available—we can bootstrap robust statistics like the median as easily as the mean.

For the promise of simulation and bootstrapping to be realized, they must be available and easy to use in general-purpose statistical software, complete with the exploratory data analysis and inferential capabilities required in teaching and practice. We discuss some of the available software for simulation and bootstrapping, in particular software built on S-PLUS.

Key words: bootstrap, resampling, simulation, statistics education.

1 Introduction

There is an enormous opportunity to improve statistics teaching in the United States. Statistical methods are required by many disciplines, and dissatisfaction with current courses is widespread; [45] summarizes:

> Service courses in applied statistics abound in college and graduate school curricula around the world ... The clear consensus, among students and professors alike, is that too many of these applied statistics courses are far from successful (3, 9, 25, 33). In Dallal's (13) words, "the field of statistics is littered with students who are frustrated by their courses, finish with no useful skills, and are turned off to the subject for life" (p. 266). Joiner (27) gave service courses in statistics "a grade of F" for being unmitigated failures (p. 53). These courses frequently receive the worst evaluations in a school.

Students often memorize the apparent cookbook of formulas just long enough to pass tests, without gaining or retaining the understanding of statistical reasoning that is essential for many real world problems.

One major problem is that students have difficulty with fundamental concepts involving randomness. The usual approach to teaching about randomness involves substantial mathematics, which is a barrier for many students, and provides little intuition even for mathematically-sophisticated students.

Simulation and bootstrapping (SAB) methods offer a way to teach statistics and probability by using the computer to gain direct experience and intuitive understanding through graphics. These methods are conceptually simple, because they make use of the duality between a distribution and a sample from that distribution. While the underlying distribution is abstract, even a beginner easily uses histograms to visualize the distribution of a data set. Similarly, a sampling distribution—the distribution of a statistic (such as a sample median, or regression coefficient) that summarizes a set of data—is abstract, but a sample of statistic values generated via simulation is easily visualized. Familiar exploratory data analysis techniques can be applied to graphically summarize features of the sampling distribution, leading to understanding of the random behavior of statistics.

> Many of the basic ideas in probability and statistics seem exceedingly difficult for most students to grasp. ... Computer graphics are very effective in conveying an understanding of fundamental concepts through pictorial representations. For example, key ideas such as the Law of Large Numbers, random walks, and the Central Limit Theorem come alive before one's eyes through computer graphics. Direct experience and actual experimentation is the best way the student can obtain a feeling for these concepts. [42]

Figure 1 shows a typical demonstration of the Central Limit Theorem. Normal probability plots can be included to provide better visual indications of departures from normality. The value of this demon-

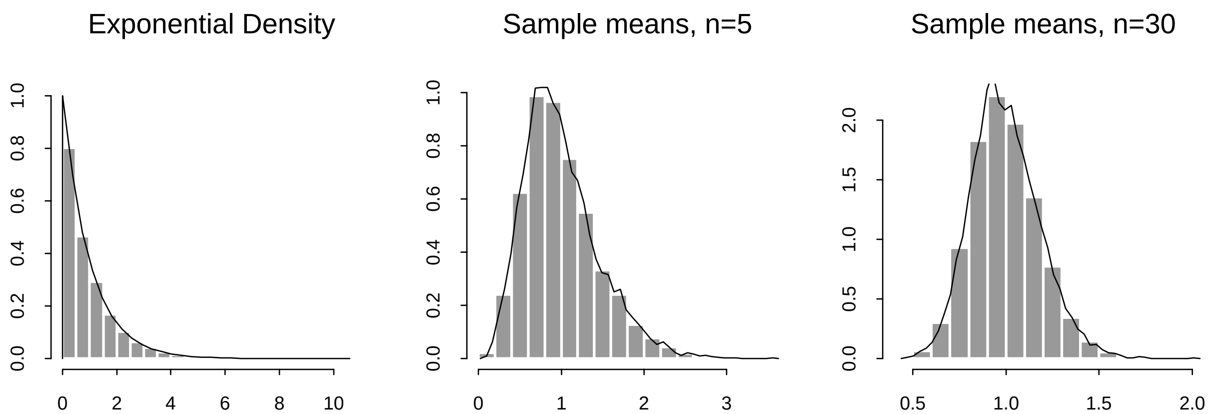

Figure 1: Investigation of Central Limit Theorem. The left panels show a histogram of a large sample from an exponential (skewed) population, with the density curve overlaid. The other panels show histograms for a random set of means of samples of size 5 and 30, respectively. The right panel is more nearly normal, demonstrating the increased accuracy of normal approximations for larger sample sizes.

stration is enhanced by putting students in control of these simulations, and letting them investigate the effects of changing sample size and skewness.

The bootstrap makes it practical to expand the statistical repertoire to make greater use of modern robust and/or nonparametric statistical methods, and to avoid unrealistic assumptions. For example, introductory statistics classes discuss medians and trimmed means, which are less sensitive to outliers than a sample mean; but these robust methods are ignored in the confidence interval and hypothesis testing chapters because the mean is easier to deal with mathematically. The bootstrap makes it just as easy to do inference for the median as the mean. Finally, even when classical methods can be used, there are often more accurate bootstrap alternatives.

The idea of using SAB for teaching is not new, but several impediments have stood in the way of widespread adoption of these methods in education: theoretical understanding of bootstrap methods, computer hardware costs, implementation in commercial software. The theory of bootstrap methods has now matured to the point that the methods are largely accepted, and computing costs continue to plummet. Some suitable software is now becoming available. In the next section we review some of the literature on the use of SAB for teaching, and give examples where it is useful. In Section 3 we discuss software.

2 Applications of Simulation and Bootstrapping

The idea of using simulation methods in teaching has engendered considerable interest over the years; [29] collected 175 references on the topic, and there are many newer articles. Writers have used simulation in a variety of ways, including simulations of random walks to help students understand the role of randomness in runs, streaks and slumps [39, 42, 8], other probability problems [41, 8], sampling distributions, the law of large numbers, and the central limit theorem [19, 42, 22, 16, 8] sampling with and without replacement [41] confidence intervals and hypothesis tests [16, 41], ANOVA and regression [16], Bayesian methods [38, 2], and the bootstrap [39, 34, 46, 44, 14].

The basic idea in simulation is to emulate real life, where one collects a sample of random data (using a survey or an experiment), and summarizes the data graphically or numerically. In simulation one generates a sample of random data on the computer in a way that mimics a real problem, and summarizes that sample in the same way. However, instead of doing this only once, one may do it many times, to investigate how such summaries vary! For example, the middle and right panels of Figure 1 show how the means of random samples vary.

Figure 2 shows results from three additional simulations. The first shows 30 student-t confidence intervals, each constructed from a sample of normal data, with the true mean shown as a vertical line; horizontal line segments that intersect the vertical correspond to confidence intervals that cover the true value. Simulations such as this one are useful in helping students understand the definition of a confidence interval ("an interval created using a procedure that covers the true value 95% of the time"). The second shows the running mean of Cauchy observations. This is useful for comparing non-robust and robust statistics, and in helping mathematical statistics students understand why the Cauchy distribution has no mean, even though it is so nicely symmetric—occasional outliers cause large jumps in the running mean, so that it cannot converge to 0. On the other hand, a trimmed mean is insensitive to those outliers and does converge quickly. The third

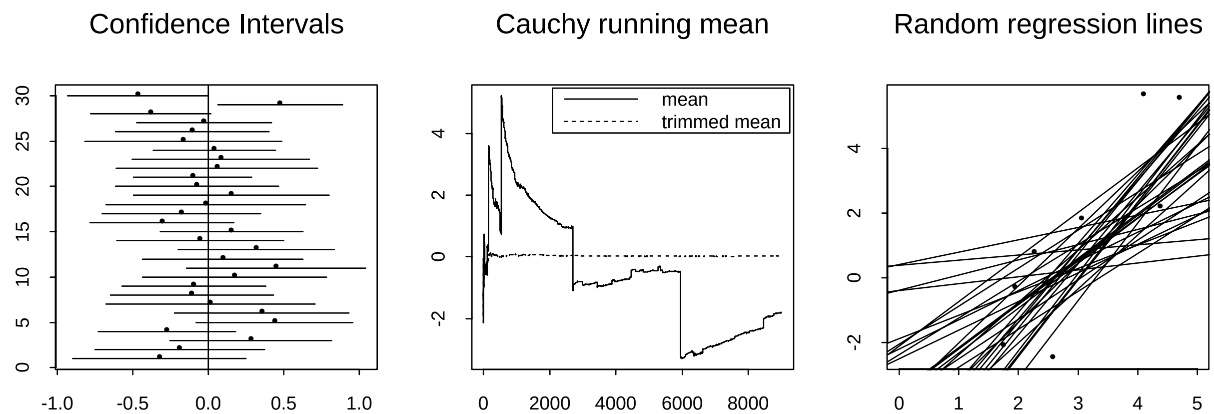

Figure 2: The left panel shows 30 95% confidence intervals. The central panel shows running means and running trimmed means for Cauchy observations. The right panel shows regression lines from random sets of data.

shows linear regression lines, each computed from a random sample of observations generated according to the standard regression model. The variation in the heights of the lines above any fixed x gives students a better feeling for the standard error of a regression prediction than they get from the usual formula $s\sqrt{1/n + (x - \overline{x})^2 / \sum (x_i - \overline{x})^2}$; the increasing variation for both large and small values of x also demonstrates why extrapolation is inaccurate. Furthermore, this should be just one of a series of simulations in which students study the effects of changing parameters – sample size, residual standard deviation, and spread of the x values – and of violations of standard assumptions – normality, homoskedasticity, fixed x values, and linear relationship.

2.1 Statistical Methods — Bootstrap

The previous examples are all "probabilistic" simulations, with samples generated from known underlying distributions. But in statistical practice, and increasingly in good data-driven introductory statistics courses, the underlying distribution is unknown and all that is available is the data. This brings us to the bootstrap. The basic idea is simple—to estimate the sampling distribution of a statistic, we can't sample from the underlying distribution because it is unknown, so instead we sample from a good estimate of it, *the observed data*.

For example, in Figure 2, the random samples for which lines are computed could as easily be drawn from the observed data as from an underlying distribution.

Figure 3 shows a diagram of the probabilistic simulation and bootstrap simulation methods for estimating sampling distributions. Where probabilistic simulation requires sampling from the underlying distribution, which is unknown in statistical practice, the bootstrap samples from the observed data. Note the parallels between these simulations; the bootstrap is useful not only in its own right, to obtain answers for any given data set, but pedagogically, to give a better understanding of phenomena associated with sampling distributions:

> Statistics is a subject of amazingly many uses and surprisingly few effective practitioners. The traditional road to statistical knowledge is blocked, for most, by a formidable wall of mathematics. Our approach here avoids that wall. The bootstrap is a computer-based method of statistical inference that can answer many real statistical questions without formulas. Our goal in this book is to arm scientists and engineers, as well as statisticians, with computational techniques that they can use to analyze and understand complicated data sets.
>
> The word "understand" is an important one in the previous sentence. This is not a statistical cookbook. We aim to give the reader a good intuitive understanding of statistical inference.
>
> One of the charms of the bootstrap is the direct appreciation it gives of variance, bias, coverage, and other probabilistic phenomena. [14]

Two examples are shown in Figure 4. First, for a sample of 50 survival times for patients given a new treatment, one might compute the 25% trimmed mean survival time as a measure of the effectiveness of the treatment. To measure how *accurate* that trimmed mean is (i.e. how much it would be likely to vary with a different set of patients), the usual bootstrap procedure is to generate "bootstrap samples" of size 50 *with replacement from the original sample*, compute the trimmed mean for each sam-

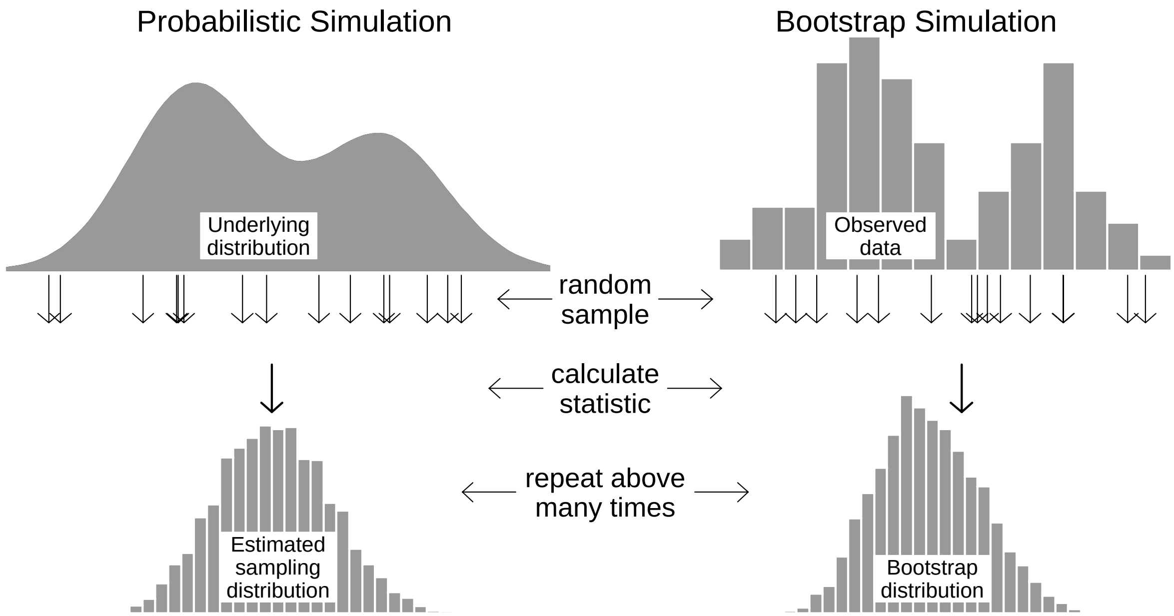

Figure 3: Diagram of probabilistic simulation and bootstrap sampling estimates of sampling distributions.

ple, and compute the standard deviation of those values. The left panel of Figure 4 shows the bootstrap distribution of trimmed means, from a set of 50 survival times of lung cancer patients from the Mayo Clinic. The bootstrap standard deviation is 52. The original trimmed mean is 338, and the average of the bootstrap samples was 340; this small difference indicates that the sample trimmed mean is a roughly unbiased estimate for the true trimmed mean (i.e. the trimmed mean of survival times of the hypothetical population of all lung cancer patients who might be given this treatment).

The second panel of Figure 4 demonstrates the concept of bias. This includes a scatterplot of 60 observations, a nonparametric regression (scatterplot smooth) through those data, 20 curves from bootstrap samples from those data, and the average of 100 bootstrap curves. Note both the variance in the bootstrap curves, and the bias – the average of the bootstrap curves is higher than the original curve in some places, and lower in others. This indicates bias in the smoothing procedure, which tends to fill valleys and level off hills. Changing a tuning parameter for the regression procedure allows for smaller bias, but at the cost of increased variability in the curves. Graphical presentation of these curves can help students gain a better understanding of this tradeoff, particularly if they are in control of the tuning parameter. (This plot is more effective in color.)

The overwhelming sentiment in the literature is that simulation and bootstrapping (SAB) are valuable in teaching introductory students. Some controlled experiments [36] support this. Yet SAB have not yet realized their potential, particularly the bootstrap. Some reasons for this will become less important over time; for example, computers are becoming cheaper, more powerful, and easier to use, faculty and students are more familiar with computers, computer labs are more common, and faculty are becoming more familiar with bootstrap methods; this last trend will accelerate as young faculty who have been exposed to the bootstrap in their studies begin teaching (many of the books on the bootstrap and other resampling methods are quite recent, including [11, 14, 15, 18, 21, 31, 40, 26, 25, 33, 35, 43]).

2.2 Other resampling methods

Many of the statistical and pedagogical advantages of the bootstrap hold for some other resampling methods, notably permutation tests. They also use simulation (or exact calculations) to produce estimates of the sampling distribution of a statistic, and are not limited to simple statistics. Permuation tests are less general than the bootstrap.

The jackknife, on the other hand, does not produce a straightforward estimate of a sampling distribution, and is less useful pedagogically.

2.3 Bayesian Methods

Sampling also has promise in the teaching of of Bayesian statistical methods; for example, the Bayesian sampling/importance-resampling (SIR) method

> provides a simple pedagogic tool for illustrating the sequential Bayesian learning process, as well as the increasing concentration of the posterior as the amount of data increases. In addi-

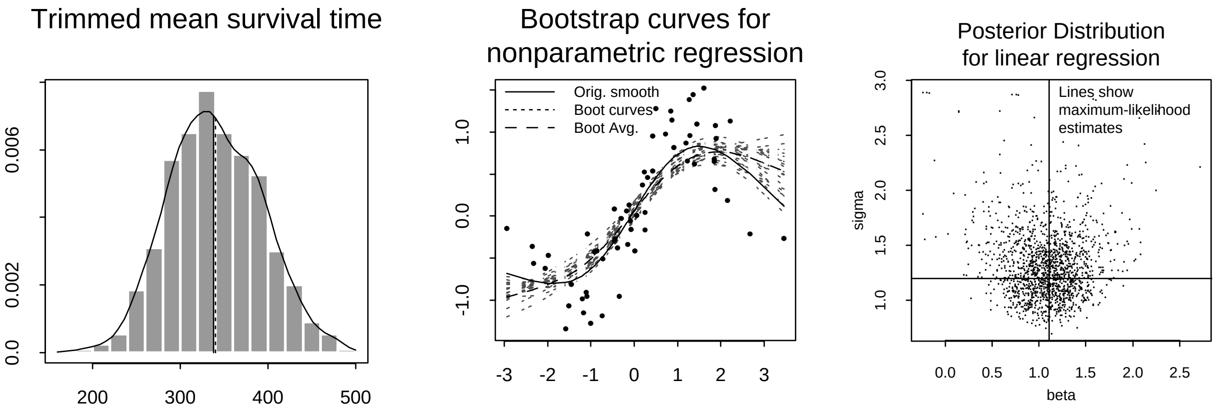

Figure 4: The left panel shows a bootstrap estimate of the distribution of the 25%-trimmed mean for survival time. The center panel shows a set of data, a nonparametric regression (scatterplot smooth) through the data, 20 randomly-selected bootstrap curves, and the average of 100 bootstrap curves. Differences between the original and average of the bootstrap curves suggest bias. The right panel shows a sample obtained by the sampling/importance-resampling procedure from the posterior distribution of slope and residual standard deviation parameters, for a Bayesian analysis of linear regression without an intercept.

> tion, the approach provides natural links with elementary graphical displays (e.g. histograms, stem-and-leaf displays, box-plots to summarize univariate marginal posterior distributions, scatterplots to summarize bivariate posteriors). In addition, the translation from functions to samples provides a wealth of opportunities for creative exploration of Bayesian ideas and calculations in the setting of computer graphics and exploratory data analysis (EDA) tools. [38]

This point is important—in our experience, students learn EDA techniques readily, and become comfortable with looking at sets of data. Approximating an abstract function, be it a posterior distribution or a sampling distribution, as a set of observations casts it in terms familiar to students. The right panel of Figure 4 is from a Bayesian SIR analysis of linear regression, with two parameters: intercept and residual standard deviation. The SIR sampling process produces a sample of observations from the posterior distribution, which can be used to estimate posterior probabilities or expected values, and to investigate interesting features of the posterior distribution; for example, here the plot is shaped roughly like a triangle—when the standard deviation σ is larger then the slope β is more variable. Comparing results from different sample sizes and/or prior distributions lets students investigate the effect of these inputs on the posterior distribution. In problems with more than two parameters the joint distribution of parameters can be viewed in using statistical graphical methods for viewing high-dimensional data, including brushing, rotating scatterplots, and Trellis graphics [5].

3 Software

A major cause for the delay in using SAB in teaching involves software. An apt summary of the situation is provided by [44]:

> Efron's (1982) bootstrap is an effective statistical tool that promises to have a major influence on both the practice and teaching of statistics ... Good new ideas like the bootstrap, however, often infiltrate the curriculum very slowly. One reason for this delay can be the lack of software.

Numerous authors [24, 29, 7, 1, 47, 4, 6, 32, 17, 23, 10, 27, 28] discuss software. They note that programs are

- too hard to use,
- too limited in their capabilities,
- not flexible,
- not suitable for use by students after graduation,
- too expensive, or
- not widely available.

The same software used for SAB should be used for all aspects of course, beginning with exploratory data analysis, one and two-sample tests, regression, ANOVA, etc. It should include modern statistical methods, including robust methods and smoothing and other nonparametric regression. The bootstrapping capabilities of the program should be flexible enough to handle a wide variety of statistics, provide graphical in addition to numerical output, and support not just the simplest form of bootstrapping, but also more accurate methods. For pedagogical purposes, the bootstrapping should be not a black box, but allow users to view and modify the mechanics of the routines.

Much of the available software is particularly unsuitable for bootstrap applications. This situation is beginning to change. We discuss some of the available software next.

Disclaimer—I work for MathSoft, the vendor of S-PLUS. All opinions in this report are my own and not those of MathSoft. Some of the opinions were formed while I was teaching, before working at MathSoft. I am not equally familiar with all programs discussed.

3.1 Statistical Package vs. Interactive Language

For most instructors, the choice of software for a course will come down to a choice between a statistical package, such as BMDP, DataDesk, JMP, Minitab, NCSS, SAS, SPSS, Stata, Statistica, and Systat, or an interactive high-level language such as S-PLUS, Lisp-Stat, R, Matlab, and Mathematica. For SAB, the languages win, in particular S-PLUS. Efron, inventor of the bootstrap, noted [13] that

> my bootstrapping has increased considerably with the switch to S, a modern interactive computing language. ... My guess is that the bootstrap (and other computer-intensive methods) will really come into its own only as more statisticians are freed from the constraints of batch-mentality processing systems like SAS.

(S-PLUS is built on the underlying S language). [30] adds

> The S language may continue to provide the simplest bootstrap programming in the future.

Disadvantages of packages for SAB include:

- lack of programming capabilities, in particularly looping and bootstrap sampling capabilities
- lack of communication between programming capabilities and canned statistical procedures
- closed systems (so students can't see what's happening)
- lack of flexibility (e.g. support bootstrapping for one-sample problems but not stratified sampling or multiple-sample problems),
- difficulty in doing any but the simplest bootstrap procedures.

The packages vary in these regards. The following summary is based on personal experience and conversations with the vendors.

DataDesk has a scripting language that can interact with its procedures, but the language is awkward for bootstrapping. Alternatively, a large data set containing all bootstrap samples can be created, split into individual bootstrap samples, and simple summary statistics calculated for those samples (this supports bootstrapping a mean or median, but not a two-sample statistic like a regression coefficient).

JMP can't do bootstrapping. A scripting language is under development.

Minitab offers a command language, and canned procedures that can output certain numerical statistics to the command language. It is possible to bootstrap those statistics, but not others.

SAS has a macro language that supports looping, and procedures. Some numerical statistics from the procedures are accessible to the macro language. There is a bootstrap macro available. There is also a "matrix" language available.

SPSS can't do bootstrapping.

Statistica has a "command" language containing many statistical procedures, and a "basic" language that supports looping and subscripting, but the command language can't currently be called from the basic language. Both the command and basic languages can be controlled from Visual Basic.

Systat has simple bootstrapping procedures built in at a low level to many of its procedures. The set of bootstrap statistics is saved in a vector, which the user may explore, but bootstrap analysis routines are not provided.

3.2 S-PLUS

S-PLUS is a flexible data analysis environment built on the S language originally developed at Bell Labs [9]. The design of S-PLUS is well-suited for education; [3] notes that it

> has several advantages over computer packages (such as SPSS, BMDP, and SAS) for use in statistics classes. S creates an environment in which the student can exercise creativity. ... From the teacher's standpoint, the most useful feature of S is that you can solve a problem in S by entering a sequence of commands which are quite similar to the steps used in solving the problem by hand, so that the connection between what the computer is told to do and what one would do if doing the problem manually is quite transparent.

S-PLUS is an open system; users can view and modify existing functions, and create their own functions. Professors can write functions for demonstration purposes, and students can look inside functions to see how results are obtained. Professors may choose to hide details of a simulation using graphical interfaces and high-level functions that can work like canned procedures, or may show their students the details.

S-PLUS is extensible, using functions written in the S-PLUS object-oriented language, C and Fortran. Users have posted 215 packages to statlib (see http://lib.stat.cmu.edu/S), most containing multiple functions. Two of those packages are sets of bootstrap functions that accompany the bootstrap books of [14, 11].

S-Plus includes a canned (but flexible and user-editable) `bootstrap` function that works with arbitrary user-supplied statistics, accessible via a graphical interface or by typing commands. For example, the first two commands here perform bootstrap sampling for a trimmed mean, save the results as "`BootResults`", and create the plot shown in the left panel of Figure 4:

```
BootResults <- bootstrap(lung.survival,
     mean(lung.survival, trim=.25))
plot(BootResults,
     main="Trimmed mean survival time")
summary(BootResults)
```

The `summary` command prints numerical summaries of the results, including estimated standard errors and confidence intervals. The graphical menu-driven interface to the function facilitates the sequence of creating, plotting, and summarizing the bootstrap results.

3.3 Programming languages

If a programming language is used, it is is essential in education that it be a *high-level* language to minimize class time spent on programming [17]. Fortran, C, and Basic are low-level languages.

High-level languages intended for statistical use include S-Plus, Lisp-Stat, and R. I haven't used Lisp-Stat. R imitates S-Plus in many ways.

Matlab and Mathematica are high-level languages but have limited statistical functionality and designs which are not oriented for statistics. I have found them frustrating for statistical work, particularly with missing data.

3.4 Resampling software

Resampling Stats is a special purpose simulation and resampling program. that has been heavily promoted for teaching statistics using SAB [8, 37, 35, 34]. It is inadequate for teaching statistics. Its capabilities are extremely limited; it does not even do scatterplots, for instance. I believe Hinkley was thinking of this program when he wrote "We are also beginning to see the first wave of software products which claim to do bootstrap analysis: some of these are embarrassingly naive." [20]. [7] give a critical discussion of this program and the claims made for its value.

LogXact and StatXact do permutation tests. I have not used them.

3.5 Spreadsheets

Spreadsheets can be used for simulations [44], but I find them awkward for that purpose. Excel has serious deficiencies as a statistical package [28].

3.6 "Educational" software

Specialized statistical educational programs such as Dynamic Statistics, EESEE, DASL, ActiveStats, and Educational Companion: Statistics have various pedagogical strengths. Some of the programs have nice canned demos involving simulation, often with animation. They are not adequate for all aspects of a course, in particular for bootstrapping, but are useful in combination with other software.

4 Pedagogical Questions

We raise four issues here. The first is how much should instructors or software do for the students. An instructor may use graphical interfaces and high-level functions to hide many details of computations from students, but some students will learn more by programming simulations themselves (in a high-level language, of course). For instance, a simple program such as the following:

```
n <- length(lung.survival) # Sample size
Result <- vector("numeric", 1000)
for(i in 1:1000){
  bootSample <- sample(lung.survival,
                       size=n, replace=T)
  Result[i] <- mean(bootSample, trim=.25)
}
```

reinforces how bootstrapping works—bootstrap samples are obtained by sampling with replacement from the data, then the trimmed mean or another statistic is computed.

The second issue is what bootstrap procedures to use. There are a wide variety of bootstrap procedures. The simple nonparametric bootstrap is adequate for estimating standard errors and obtaining rough confidence intervals, and are best for initial pedagogical purposes. However, more accurate procedures are available, such as the BC-a confidence intervals [12]. These are more complicated to program. Similarly, some variance reduction procedures can dramatically improve the simulation accuracy of the bootstrap, for a given number of replications, but are more complicated to program.

I recommend that students use the more accurate bootstrap procedures by the end of the course, to improve the chances that they will use them in later practice, and because the differences between these intervals and the simplest bootstrap intervals provide an opportunity for students to better understand the effect of bias on inferences. I do not recommend that students program these themselves (except in advanced courses), but they instead be provided in easy-to-use software (such as the S-Plus `bootstrap` procedure).

The third issue is that bootstrap procedures do not work for all problems. I recommend the use of graphical diagnostics procedures [11] to help students judge whether their results are trustworthy.

The fourth point is that when using simulation methods to visualize sampling distributions, the number of replications should be large. With fewer replications there are larger random fluctuations in the estimated sampling distribution, e.g. multiple modes in an estimated density curve, and these may confuse students. They may have trouble distinguishing randomness due to random selection of data from randomness due to using small numbers of replications.

Acknowledgments
I wish to thanks Alan Li, Jagrata Minardi, David S. Moore, and James Schimert for helpful comments.

References

[1] J. Albert and M. Berliner. Review of "Resampling Stats (Version 3.14)". *The American Statistician*, 48:129–131, 1994.

[2] J. H. Albert. Teaching Bayesian statistics using sampling methods and Minitab. *The American Statistician*, 47:182–191, 1993.

[3] D. R. Barr. Use of S in statistics courses. In *Proceedings of the Section on Statistical Education*, pages 134–139. American Statistical Association, 1985.

[4] R. Baxter, N. Fisher, and M. Walmsley. A graphical user interface for S. In H. J. Newton, editor, *Computer Science and Statistics: Proceedings of the 24th Symposium on the Interface*, volume 24, pages 359–365, 1992.

[5] W. Cleveland Becker, R. and L. Clark. Trellis graphics. http://netlib.bell-labs.com/cm/ms/departments/sia/project/trellis.

[6] J. M. Bland. Using a microcomputer as a visual aid in the teaching of statistics. *The Statistician*, 33:253–259, 1984.

[7] A. Boomsma and I. W. Molenaar. Resampling with more care (with discussion). *Chance, New Directions for Statistics and Computing*, 4(4):25–31, 1991.

[8] P. C. Bruce. Resampling as a complement to *against all odds*. In *Proceedings of the Section on Statistical Education*, pages 85–91. American Statistical Association, 1992.

[9] J. M. Chambers and T. J. Hastie. *Statistical Models in S*. Wadsworth, California, 1992.

[10] P. B. Collings. Computer usage in introductory statistics: the good and bad. In *Proceedings of the Section on Statistical Education*, pages 1–6. American Statistical Association, 1992.

[11] A. Davison and D. Hinkley. *The Bootstrap*. Cambridge University Press, 1997. in press.

[12] B. Efron. Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, 82:171–200, 1987.

[13] B. Efron. discussion of "Bootstrap: More than a stab in the dark?". *Statistical Science*, 9:396–398, 1994.

[14] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.

[15] P. Good. *Permutation Tests: a practical guide for testing hypotheses*. Springer-Verlag, 1994.

[16] F. Gordon and S. Gordon. Sampling + Simulation = Statistical Understanding: computer graphics simulations of sampling distributions. In F. Gordon and S. Gordon, editors, *Statistics for the Twenty-First Century*, MAA Notes No. 26, pages 207–216, Washington, D.C., 1992. Mathematical Association of America.

[17] M. A. Greene. Computing with Hogg and Craig. In *Proceedings of the Section on Statistical Education*, pages 45–53. American Statistical Association, 1991.

[18] P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer, New York, 1992.

[19] D. K. Hildebrand and P. J. Hildebrand. Simulation studies for a serious service course. In *Proceedings of the Section on Statistical Education*, pages 135–139. American Statistical Association, 1986.

[20] D. Hinkley. discussion of "Bootstrap: More than a stab in the dark?". *Statistical Science*, 9:400–403, 1994.

[21] J. S. U. Hjorth. *Computer intensive statistical methods: Validation model selection and bootstrap*. Chapman and Hall, London, 1994.

[22] T. Kersten. Computer simulations to clarify key ideas of statistics. *Two-year College Mathematics Journal*, 14:416–421, 1983.

[23] D. S. Kleiner and M. Borenstein. PC-based simulation of statistical power with business law applications. In *Proceedings of the Section on Statistical Education*, pages 210–213. American Statistical Association, 1993.

[24] H. Krieger and J. Pinter-Lucke. Computer graphics and simulations in teaching statistics. In F. Gordon and S. Gordon, editors, *Statistics for the Twenty-First Century*, MAA Notes No. 26, pages 198–206, Washington, D.C., 1992. Mathematical Association of America.

[25] E. Mamman. *When does bootstrap work? Asymptotic results and simulations*. Springer Verlag, 1992.

[26] B. F. J. Manly. *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London, 1991.

[27] R. G. Marks. Emphasizing concepts in graduate statistics service courses. In *Proceedings of the Section on Statistical Education*, pages 7–11. American Statistical Association, 1992.

[28] J. D. McKenzie and W. H. Rybolt. What is the most appropriate software for a statistics course? In J. Sall and A. Lehman, editors, *Computer Science and Statistics: Proceedings of the 26th Symposium on the Interface*, volume 26, pages 380–384, 1994.

[29] J. D. Jr. McKenzie. Why aren't computers used more in our courses? In *Proceedings of the*

Section on Statistical Education, pages 12–17. American Statistical Association, 1992.

[30] M. P. Meredith and J. G. Morel. discussion of "Bootstrap: More than a stab in the dark?". *Statistical Science*, 9:404–406, 1994.

[31] E.W. Noreen. *Computer Intensive Methods for Testing Hypotheses: An Introduction*. Wiley, New York, 1989.

[32] J. L. Romeu. Statistical education via simulation. In *Proceedings of the Section on Statistical Education*, pages 205–210. American Statistical Association, 1995.

[33] J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer-Verlag, New York, 1995.

[34] J. L. Simon. Resampling and the ability to do statistics. In *Proceedings of the Section on Statistical Education*, pages 78–84. American Statistical Association, 1992.

[35] J. L. Simon. *Resampling: The new statistics*. Resampling Stats, Arlington, Virginia, 1992.

[36] J. L. Simon, D. T. Atkinson, and C. Shevokas. Probability and statistics: Experimental results of a radically different teaching method. *American Mathematical Monthly*, 83:733–739, 1976.

[37] J. L. Simon and P. Bruce. Resampling: a tool for everyday statistical work. *Chance: New Directions for Statistics and Computing*, 4:22–32, 1991.

[38] A. F. M. Smith and A. E. Gelfand. Bayesian statistics without tears: A sampling – a resampling perspective. *The American Statistician*, 46:84–88, 1992.

[39] J. L. Snell and W. P. Peterson. Does the computer help us understand statistics. In F. Gordon and S. Gordon, editors, *Statistics for the Twenty-First Century*, MAA Notes No. 26, pages 167–188, Washington, D.C., 1992. Mathematical Association of America.

[40] P. Sprent. *Applied nonparametric statistical methods*. Chapman and Hall, London, 1989.

[41] E. A. Tanis. Computer simulations to motivate understanding. In F. Gordon and S. Gordon, editors, *Statistics for the Twenty-First Century*, MAA Notes No. 26, pages 217–225, Washington, D.C., 1992. Mathematical Association of America.

[42] H. Tijms. Exploring probability and statistics using computer graphics. In F. Gordon and S. Gordon, editors, *Statistics for the Twenty-First Century*, MAA Notes No. 26, pages 189–197, Washington, D.C., 1992. Mathematical Association of America.

[43] P.H. Westfall and S.S. Young. *Resampling–Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley, New York, 1993.

[44] T. R. Willemain. Bootstrap on a shoestring: Resampling using spreadsheets. *The American Statistician*, 48:40–42, 1994.

[45] J. B. Willett and J. D. Singer. Providing a statistical "model": Teaching applied statistics using real-world data. In F. Gordon and S. Gordon, editors, *Statistics for the Twenty-first Century*, MAA Notes No. 26, pages 83–98, Washington, D.C., 1992. Mathematical Association of America.

[46] T. H. Wonnacott. More foolproof teaching using resampling. In *Proceedings of the Statistical Education Section*, pages 273–274, Washington, D.C., 1992. American Statistical Association.

[47] F. W Young, R. A. Faldowski, and M. M. McFarlane. ViSta: A visual statistics research and development testbed. In H. J. Newton, editor, *Computer Science and Statistics: Proceedings of the 24th Symposium on the Interface*, volume 24, pages 224–233, 1992.