Journal
Technology Innovations in Statistics Education, 5(1)

Author
Carver, Robert H

Publication Date
2011-06-14

DOI
10.5070/T551000043

1. INTRODUCTION

In a recent rethinking of our introductory courses, George Cobb reminds us just how pervasive have been the accommodations to problems of computability. Many of our core topics are progressively complex adjustments to approximations that we rely on because more direct methods were not previously practically computable. The normal distribution dominates introductory courses as we lead students progressively further afield from the essence of statistical thinking. But now we have cheap and powerful computing that permits us to remove the normal distribution from the center of our universe and instead replace it with the fundamental logic of inference. Liberated from the "tyranny of computability" we can "emphasize the 3R's of inference: Randomize, Repeat, Reject any model that puts your data in its tail" (Cobb, 2007). We need not worship at the altar of the Gaussian model, but rather we can leverage computing power to reimagine the content and structure of introductory statistics courses.

In statistics education we have a long history of questioning what we should teach, when we should teach it, and how we should teach it (Federer, 1978; Hogg, 1991; Meng, 2009; Moore, Cobb, Garfield, & Meeker, 1995; Vere-Jones, 1995). There is an equally rich history of adapting to new technologies, from calculators to computers (Friedman & Stuetzle, 2002; Phillips, 2001; Schatzoff, 1968; Tukey, 1972). This paper takes up Cobb's exhortation and offers a recipe for a course in the spirit of his article. The recipe anticipates local variation according to the needs and realities of readers' home institutions, but should be adaptable by many readers.

In 1937 Robert Cobb introduced a salad at the Brown Derby restaurant in Hollywood, California featuring an assortment of vegetables, meat, eggs and cheese with each ingredient chopped and artistically arranged in adjacent mounds on a dinner plate. A Cobb salad is a full meal in itself—rich in nutrients and flavor. It is loaded with proteins and fats providing the diner with a substantial meal of bite-sized elements, unified with a savory dressing. Though not precisely in keeping with the latest nutritional guidelines on healthful eating, it does have quite a lot to recommend it: with some subtle modification, it provides the diner with a balanced variety of protein, fats and carbohydrates, as well as an appealing mix of textures, colors and flavors.

Our standard emphasis on technique has too often produced introductory statistics courses that are green salads: an enormous amount of space on the plate is occupied by bulky greens that have relatively little flavor or nutritional value, take a long time to chew, and ultimately leave the diner unsatisfied. Occasionally, there are some pickles included: old, reworked, or artificial datasets that appeal only to particular tastes or, worse still, lie limply on the plate. The recipe outlined here reduces some of the space-filling and watery leafiness and adds substantial ingredients that will fortify students and remain with them for a long time.

2. A FRAMEWORK: CONSTRAINED OPTIMIZATION

We begin building the course within a framework of multi-objective constrained optimization. The animating idea in Cobb's article is that an ancient constraint has been lifted; we ought to productively think about our objective function(s) and our constraints. What are we optimizing, and what are the critical constraints?

2.1 What Shall We Optimize?

We live in an age when statistical thinking is one major modality of truth-seeking, policy development and decision-making. Scholars from various nations and cultures have cited the idea of "statistics for citizenship" for years (Bartholomew, 1995; Gal, 2003; Utts, 2003; Wild, 1994). If we are to educate for world citizenship, we also must pursue the objective of statistical thinking**;** indeed our literature often points to it as the objective *par excellence*. One recent discussion notes that

"Statistical thinking uses probabilistic descriptions of variability in (1) inductive reasoning and (2) analysis of procedures for data collection, prediction, and scientific inference" (Brown & Kass, 2009). The authors go on to note that "Statistical models of regularity and variability in data may be used to express knowledge and uncertainty about a signal in the presence of noise, via inductive reasoning." I suggest that we emphasize the logic of inference in order to maximize our students' preparedness for world citizenship, though the specifics of that preparation will have quite wide variation across nations.

If we define an objective function to represent statistical thinking, then clearly one important term in that function should be the "logic of inference" as Cobb says. Other factors that contribute to statistical thinking would include understanding important foundational concepts such as populations, samples, variability, data, and description. Each portion of Section 3 below focuses on one or more of the specific terms in the objective function: (1) populations and variability, (2) samples and variability, (3) the logic of inferential modeling, and (4) applications of inference.

It is probably less important to agree on a single universal objective or set of objectives than it is to have clarity about the core objectives within the design of one university's course. For the sake of this paper, let's consider statistical thinking to be the prime objective. Suitable secondary goals include citizenship and global awareness, cultivation of interest in quantitative modeling, facility with statistical computing, and awareness of quantitative skills as valuable in the job market.

2.2 Which Constraints Matter?

Those who teach an introductory statistics course are well aware of the typical constraints that impinge on the design and conduct of the course. Cobb's insight, of course, is that some of the long-standing constraints are no longer relevant. Naturally the constraints vary across nations, academic programs and schools. Some are so embedded in the design of textbooks and courses that we have forgotten about them. Figure 1 identifies the key customary constraints and the associated characteristics of traditional introductory courses.

Which constraints still matter? This section briefly outlines the proposed constraint set:

- **Important problems**: Statistical thinking is brought to bear on problems of great importance and therefore illustrative problems and cases presented in our courses should be real and important. We would do well to take our lead from Hans Rosling (Rosling, Rönnlund, & Rosling, 2004) and the work of www.gapminder.org to excite students about global inequality with dazzling visual tools.

- **Best practices:** Follow Guidelines for Assessment and Instruction in Statistics Education GAISE (or comparable) guidelines.

Constraints	**Course and Textbook Characteristics**
• Assumptions about the Canon • Subject-oriented discipline requirements (psychology, engineering, business, etc.) • 14 weeks • Student backgrounds in mathematics and subject domains • Availability of computational technologies: slide rules, calculators, computers, tables (z, t, F…), applets • Availability and accessibility of real, important data sets (very limited until recently) • Class sizes & management of assignments	• Small samples and small-ish samples ($n < 20$ and $30 < n < 100$) • Artificial data and non-random samples • Instructional focus on computation • Focus on approximations to smooth distributions • Students demonstrate "skills" that are exam-friendly (finding numerical results) • Emphasis on "Mechanics" • Interpretation of significance test results

Figure 1: Traditional Constraints and Course Attributes

- **One semester**: the introductory course is often the terminal course, and we have roughly one quarter of the year at our disposal (in the U.S. a 14-week course is typical, but some are shorter).

- **Infrastructure**: Every course operates within a campus IT infrastructure and physical classrooms, and those constraints clearly have impact in this discussion. Additionally, we must recognize national differences in statistical infrastructure and student preparation in secondary in schools (Peres, Morettin, & Narula, 1985).

- **Barriers to reform**: We must also recognize the social, political, economic hurdles to curricular reform (Kerr, 1989; Obanya, 1995). These forces exist at the institutional level (campus cultures and budgets come to mind) as well as at the regional and national levels.

- **SOS disciplines**: Many courses serve the needs of programs in engineering, social sciences, business, or natural science. At institutions where service courses are common, we must meet the needs of the disciplines with whom we cooperate in offering *Subject-Oriented Statistics* (SOS) courses (Altman & Bland, 1991; Love & Hildebrand, 2002; McAlevey, Everett, & Sullivan, 2001; Meng, 2009; Smeeton, 1997; Yasar & Landau, 2003).

- **Non-negativity:** Here we should be inspired by the Hippocratic Oath: do no harm. Better still, keep the affect positive, so that on net we offer "Happy Courses" (Meng, 2009).

Cobb's original point is that the long-standing constraints of computability have been substantially relaxed or effectively dropped. Sampling distributions that presented insurmountable computational burdens can be quickly simulated (with animation, no less) and in some cases exhaustively generated by standard statistical software and cheap laptops. While the Central Limit Theorem is still an independently valuable and impressive result[1] it is no longer indispensible to either the understanding or the practice of inferential reasoning.

It's not just the computability constraints that have relaxed. We should note that in many nations we no longer should assume an absence of prior statistical background knowledge (Holmes, 2003) as secondary school students regularly have formally studied statistics. Concurrently, the web has made available vast amounts of public-domain data and there are now numerous libraries of fair-use experimental data and public domain microdata from reputable surveys. Though it varies across schools and countries, student access to computing power and appropriate software is far less a problem than just a few years ago. In some schools or nations, limited access may continue to present insurmountable barriers to implementing the ideas presented below. The expectation is that over time costs will continue to fall and access will continue to increase.

2.3 Relaxed Constraints = Opportunities

One of Cobb's messages is that defunct constraints live on in the topics we include in our introductory courses. With a new set of constraints there are a number of topics that we leave behind, making room for greater depth or for a reconsideration of those topics that we consider too advanced for an introductory course. Figure 2 lists topics and techniques which should be tossed onto the compost heap as we prepare this new salad as well as those we might include selectively to enrich the offering.

OUT: Topics We Can Afford to Drop	**IN: Formerly "Advanced" or Neglected Topics**
• Most manual computations using small datasets • All software-based computation using artificial data	• Data preparation, missing data, and data management • Writing about statistical investigations

[1] Probably more impressive to statistics teachers than to our pupils.

• Defining histogram bins • Most elementary probability • Reading tables of *t*, *z*, *F*, etc. • Normal approximations to other distributions • Z-intervals and significance tests	• Resampling and Permutation tests • Analysis of Probabilistic non-SRS data using sampling weights • Nonparametric methods • Non-linear models • Multivariate models

Figure 2: What's Out and What's In

3. AN OUTLINE FOR A 14-WEEK COURSE

Much as we think of data as a mixture of signal and noise we may do well to reexamine our course outlines similarly with an eye to improving the signal-to-noise ratio. Activities and lessons that advance statistical thinking and the logic of inference are signal, as is any content that satisfies SOS-related applications or other constraints. All other potential course elements are noise.

The course design should follow principles of good study design, following the essential structure of a Plan-Do-Report (P-D-R) cycle (Sharpe, De Veaux, & Velleman, 2010), and iterating the cycle several times within the course. In this particular proposed course, there are four iterations of the P-D-R pattern. **Planning** involves raising theory- and data-driven questions, specifying variables, and planning for data collection. **Doing** is about methods and techniques. This course proposal presents methods that satisfy constraints and/or increase facility with the logic of inference and statistical thinking. For each "doing" segment, a micro P-D-R cycle will repeat according to the conventions of the technique. **Reporting** focuses on resolution of those important problems, assessing the extent to which conclusions can be drawn, with direct attention to language suited to communicating within our allied disciplines (Radke-Sharpe, 1991; Samsa & Oddone, 1994; Wild, 1994).

In keeping with the Cobb salad metaphor, let's treat the four P-D-R cycles as four nutritious elements arranged artfully in *mounds*, and describe the topics that form ingredients for each mound. Instructors will likely have favorite data sets and illustrative examples. The following discussion presents the topics treated in each cycle, as well as some illustrative examples. The purpose is to invite readers to reconsider how and what we teach in the light of technologies at our disposal, keeping in mind that statistical software obviates the need for some traditionally critical topics.

Readers will note that the specificity of examples diminishes as the discussion proceeds through the four mounds. This is based on the author's view that earlier mounds are applicable to a wide variety of educational settings, and therefore will suit the needs of most instructors. Conversely, the that later mounds will tend to vary more widely by institution, so course designers and instructors will want to develop their own examples. In addition there are probably diminishing returns to repeating extended illustrations.

3.1 Course Structure and Context

As noted earlier, the primary assumed objective is to maximize students' facility with statistical thinking. Additional objectives include generating genuine engagement with the discipline and developing a positive affect about statistics. Before presenting the proposed course approach, it should be noted that the author has conceived of this course in a setting of classes of 25 to 35 students, each with in-class access to computers. That said, though, the approach should be adaptable to larger classes in which an instructor demonstrates the software, and then students complete structured assignments outside of class either in a lab setting or using their own computers.

My goal here is to spur practical and creative thinking about how to change the introductory course as we continue to apply computing power to the big ideas of statistical thinking. Traditionally we devote considerable resources to calculations associated with (for example) theoretical distributions and there is no longer a need to do so. This article does not layout a turnkey plan for a course, but does provide a framework and some direction for working through the details of implementation.

We also should not underestimate the potentially constraining effects of campus cultures and conventions. A course like the one described here invites students to collaborate, to work independently outside of class, and to raise questions guided by their data. The approaches suggested will more readily take root in some settings than in others.

3.2 A Word About Software

Because available computing power is a central driver in the argument here, it is natural to build the proposed framework using current software. I believe that this proposed course works best when students have the software installed on their own computers, and that is likely to be a problem at many institutions. Ideally, schools have licensing arrangements with the relevant vendor that allow faculty and students to download and run the software on their own computers, permitting maximum access for data exploration and tinkering.

My bias is towards packages written specifically for statistical analysis with sizeable user bases in industry and academia (Minitab, SPSS, Stata, SAS, JMP, R, etc.) rather than widely available spreadsheet software or even marvelous tools like Fathom or Trendalyzer. This is not to denigrate the alternatives but rather to introduce students to tools that will serve them well in graduate programs or in the job market while still facilitating the habits of statistical thinking. The illustrations below use JMP in part because I find JMP to be an excellent environment for undergraduate introductory statistics students: it is menu driven, native to both Macintosh and PC operating systems, thoroughly visual, and consistently links graphics to the analytical methods. JMP does have an extensive scripting language but with menus the cognitive barrier to entry is quite low, permitting instructors to devote energy to teaching statistical thinking rather than teaching programming.

JMP is not unique in these respects, though its ease of use, visual approach and other advantages certainly make it very attractive for this purpose. This course could also be implemented with most of the other programs mentioned, thought the burdens on students might be greater.[2]

3.3 Mound 1: Populations and Variability (4 weeks):

The course should start with a vivid presentation of an important international problem relevant to a SOS discipline. Rich areas for problem selection include social justice, public health, or quality of life (Lesser, 2007; Rosling, et al., 2004). This primary objective in this mound is to lay the foundations of the core concepts of variability within populations or processes and the generation of data about a population. The secondary objectives here (secondary to those of developing statistical thinking) include making a clear connection between a SOS discipline and statistics, as well as introducing students to the software environment. The constraints are those already discussed in Section 2.2 and 2.3. This illustration serves several SOS disciplines, including economics, political science, sociology, or any of the health professions. It is also accessible to most undergraduates.

As we'll soon see, the illustration included here is about the relationship between fertility and wealth around the world. Birth rates, life expectancies, family sizes, poverty and wealth inequality underlie a huge number of social, economic and political struggles. How do we study and describe the dimensions of such problems? What do we mean by variation and to what extent is variation at the root of such problems? How do we measure and characterize variation (concepts, constructs,

[2] One might be hard-pressed to make this approach work with spreadsheets, unless supplemented with many add-ins.

measures)? Here the students first meet concepts of study design for experimental, observational, survey research.

The first P-D-R cycle presented below runs as follows. Through lecture and demonstration, the instructor explains a *plan* to access data published by the U.N., focusing on two variables that relate to the fertility-wealth dynamic. Once the data set is assembled, the plan calls for creating some graphs to explore and describe the variation of each variable singly and also to look at their joint variation. Once that plan is clarified, either in the classroom or through a structured lab or homework assignment, students *do* create the graphs (or in one case manipulate an instructor-created graph). Finally, students *report* on what they did, what they saw, and what they conclude about the univariate and bivariate variation.

This mound presents one population dataset and introduces data sources, discussing their credibility and reliability. To illustrate the kind of examples and analyses in this portion of the course, I use the Millennium Development Goals Indicators from the United Nations("Millennium Development Goals Indicators," 2010). These data feature prominently on the gapminder.org site. Before any exploratory analysis, it pays to ask the class to speculate about the origins of such data: how does the U.N. know each nation's fertility rate, for example? Should we trust the numbers? Do all countries have equally rigorous systems in place to record and aggregate such figures? Are all Internet data sources equally credible?

In JMP we can demonstrate simple database queries and data management concepts, opening the door to enormous databases typically available to governments and industry. We use JMP's visual interface to develop deeper understanding of data types, of distributions and density, and how graphs and summary measures represent distributions. We can also clarify the conceptual notion of a population as dynamic, subject to variability over time as well as the concept of unit of analysis.

The first set of examples focus on the Fertility rates and Gross Domestic Product of each nation in 2005[3], and we inquire about the connection, if any, between fertility rates and national wealth. Like many software products, JMP's dialog boxes reinforce some key ideas (*e.g.*, using icons to represent data types) and anticipate analytical options, as shown in Figure 3.

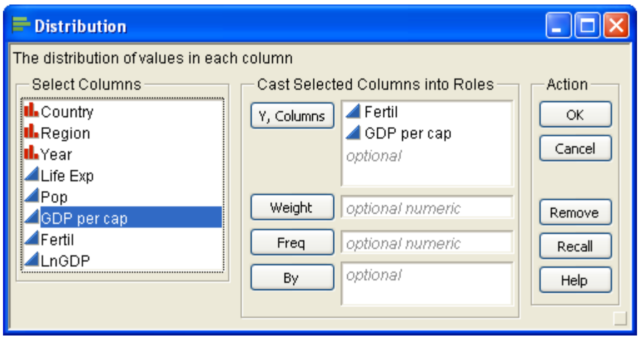


Figure 3: A JMP Dialog to Summarize a Distribution

This dialog generates the output displayed in Figure 4, combining graphical summaries with the standard set of summary measures. Moreover, it invites conversation about the shape of these two distributions, about why variables take on characteristic shapes, about the effect of skewness on measures of center, and about the utility of transformations such as the natural log. It this example we might also usefully discuss missing data: the UN reports fertility rates from 189 countries, but GDP per capita for only 157. At an early point in the course students make inquiries about data collection, integrity and statistical bias, as well as think about reasonable ways to cope with missing data.

[3] The full dataset used in this example contains national figures for every five-year interval from 1970 through 2005.

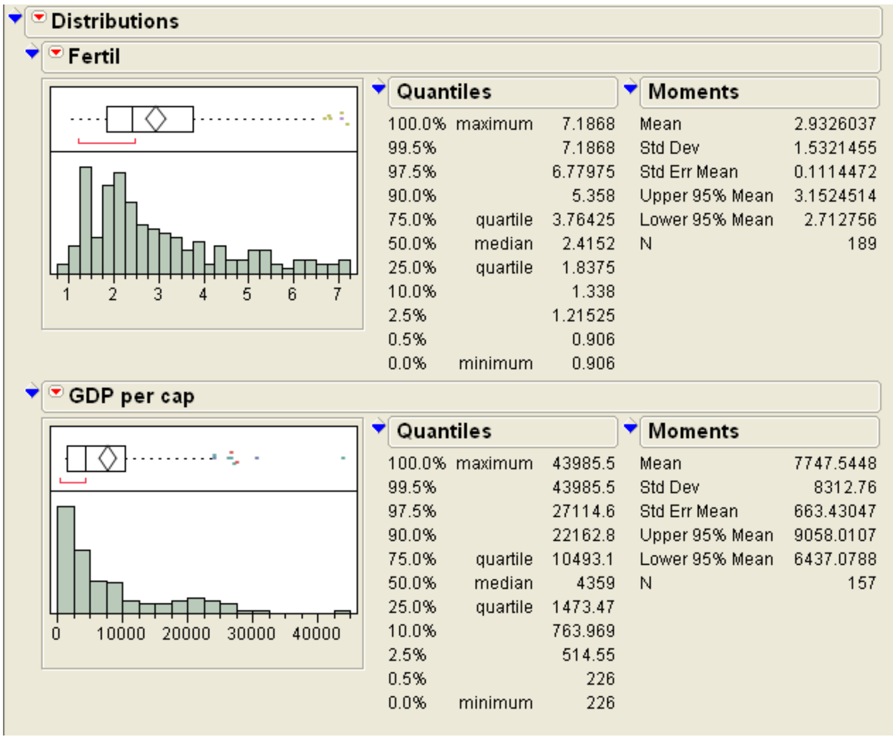


Figure 4: Two Distributions

Interactive software lends itself to active learning. Rather than becoming enmeshed in the computational minutiae of bin width specification, students manipulate the default bin widths in a JMP histogram and see how the shape of the histogram changes; in the limit, they request a *Shadowgram* which visually averages the effects of many possible bin widths, creating a very natural bridge between discrete and continuous data. (See Figure 5). In the left panel of the figure the user slides the grabber tool from left to right adjusting bin widths interactively.

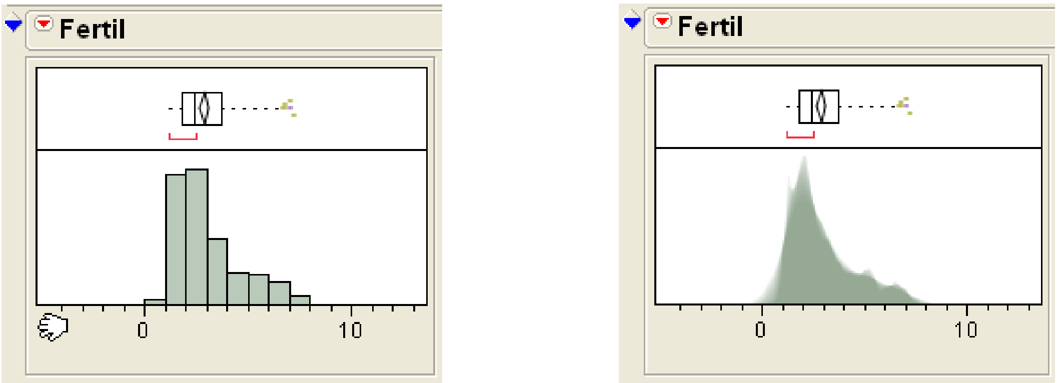


Figure 5: Two More Visualizations of the Fertility Data from Figure 4

Why have we traditionally taught students how to calculate histogram bin widths? Might it be because that is a necessary step when the available technology was graph paper? If so, why do so many statistics texts continue to include this topic? This is a small but instructive example of ways in which some course topics persist only to satisfy defunct constraints. It may be simpler to teach and test proper construction of bins than to do so for thinking about what a histogram reveals, but if the course objectives are about statistical thinking, then the choice of topics and of assessments should also be about thinking.

In a course free of the constraint of limited computational power, assignments and exams should focus on the higher-level outcomes associated with objectives, such as the interpretation of graphics rather than their creation. Homework assignments should pose questions that require some software-based analysis, followed by short-answer responses. Similarly, (depending on the technology infrastructure constraints on campus) exams include a combination of hands-on, real-time problem solving as well as the conventional array of question types that aim directly at the specific learning objectives of the unit.

In mound one, we use standard graphs to build students' statistical literacy skills. Software affords many opportunities for creating, modifying, and customizing graphs so that users become increasingly self-assured interpreting graphs they have created. Figure 6, for example, shows a modified boxplot of the same fertility data superimposed on all 189 observations (which are colored and marked by regions of the world). With this plot, the student begins to understand the ways in which the boxplot divides the data set into four groups and identifies outliers. It also visually suggests the presence of regional variation in fertility rates. The purple dots from Sub-Saharan Africa tend to reflect much higher fertility rates than the blue dots from Europe and Central Asia.

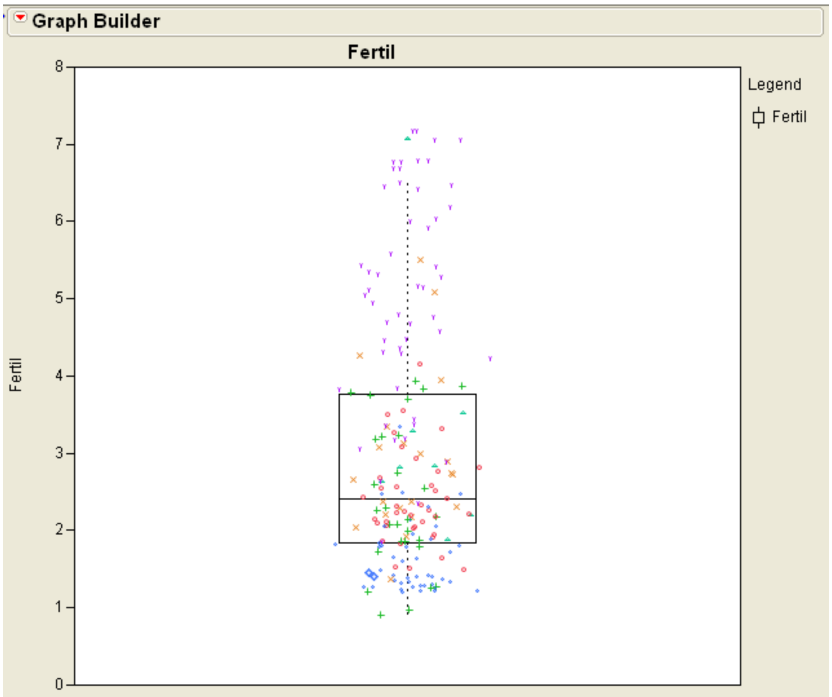


Figure 6: An Alternative Boxplot

Finally, the obvious graphical starting point for the relationship between wealth and fertility is a simple scatterplot. However, if we expand the view of the dataset to include additional years, we use JMP's bubble plot to produce animated graphs much like those generated by gapminder's Trendalyzer tool. We see a single static image of such a graph in Figure 7. This graph displays five variables simultaneously: log of GDP, fertility rate, population size, year, and world region, and is refreshingly easy to construct by simply assigning variables to roles within the graphing dialog. When animated, the general trend is for the points to drift downward and to the right: fewer babies per woman and more wealth per capita. Students immediately notice countries that deviate from the pattern, commenting on the dramatic increase in China's GDP and the disturbing decrease in the Congo. More subtle are the changes observable in Iraq and Rwanda during this period. Note that an interactive HTML version of this figure accompanies this article on the TISE website.

It may prove unreasonable to expect introductory undergraduates to produce this type of graph on their own, but the interactivity of the graphic renders it much like an applet. With suitable written instruction, students should be able to manipulate the variables and the animation in ways that help them gain insight into relationships involving several variables over time. Whether they create the graph in JMP, or use an instructor-provided graph (or visit gapminder.org for that matter), the important learning objectives can be achieved.

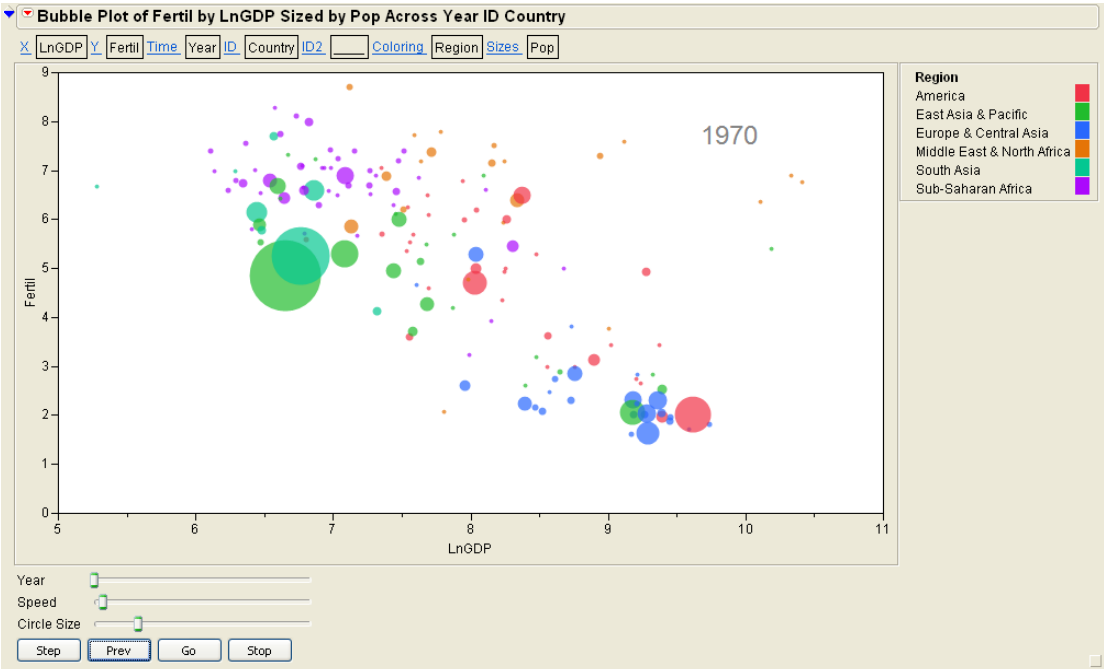


Figure 7: Bubble Plots Summarize Large Amounts of Data in Visually Gripping Ways

3.4 Mound 2: Samples and Variability (3 weeks):

Why do we sample at all? In this mound the focus shifts to element of the objective function dealing with samples, including sampling methods, sampling error, and the idea of representativeness. In this section of the course we drill deeper into descriptive methods for sample statistics. We present techniques and conventions for describing sample data, for gleaning insights and for generating hypotheses. Here, too, is the place for the foundations of experimental design and survey research (depending on SOS allies).

In this P-D-R cycle, we also focus on a secondary objective of having students experience data analysis as we practice it. With the modified constraint set allowing room for non-traditional topics, there may now be space for the unglamorous but critical processes for data cleaning, dealing with missing data, and other forms of data preparation. As appropriate, we also selectively teach how to design experiments and/or to design simple and complex samples for survey research. In this mound we introduce proper habits of speaking and writing about sample results. We reach beyond describing samples, crossing the threshold of inference even prior to a full theoretical development – to plant the seeds of the underlying concepts.

Continuing with the previous data example, when one downloads the MDG data from the United Nations, each data series forms a rectangular array: each country occupies a row and columns alternate between representing a year or a footnote. Like other popular statistical software, JMP data tables are best organized with repeated observations stacked within columns, so there is some reorganization to be done. It is not difficult, but it surely is part of the practice of statistical analysis and is teachable.

This is also the time to introduce theoretical probability distributions, moving from empirical relative frequency to theoretical models. To lay the foundations for the next section of the course and the subject of inter-sample variability, the binomial and normal distributions are likely candidates.

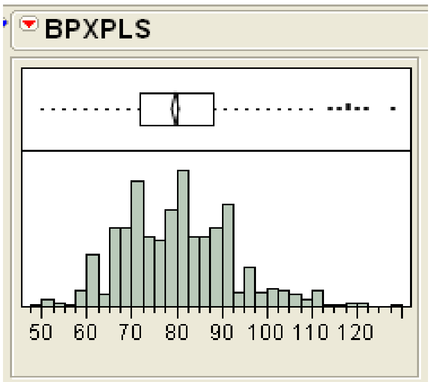
We should recognize that most of the major statistical packages have dialog boxes asking for sampling weights, and that data from widely-used survey data like the General Social Survey, the

Current Population Survey and the National Health and Nutrition Examination Survey include weighting variables. Is an introductory course too soon to introduce students in the social sciences or commerce to the need for weighting observations differentially?

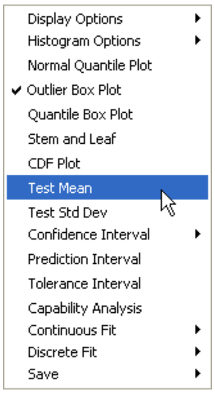
Using data from one of these familiar surveys, students develop hypotheses and then analyze the data accordingly. Within JMP there is a seamless integration of description and inference. Having generated the distribution of sample data, the context-sensitive menu displays relevant analytical options—both descriptive and inferential. To illustrate, consider the following example. At least one fitness website ("Resting Heart Rate," 2010) provides guidelines for resting heart rates depending on gender and general fitness. Heart rates typically decline with improved fitness, and vary with age (the age effect is different for men and women). For instance women between the ages of 18 and 25 in average health should expect to find their heart rates in the range of 74 to 78 beats per minute (bpm) according to the site. Women of the same age in poor condition have average resting rates above 85 bpm. But are these website guidelines credible?

To investigate the plausibility of the 74 to 78 bpm guideline we use the data from the National Health and Nutrition Examination Survey (NHANES) which is easily accessible online, subsetting the data to restrict our sample to the 619 women ages 18 to 25 years. Figure 8 illustrates the JMP analysis. There are four panels in the figure, corresponding to four distinct steps in the analysis; note that for testing purposes we use a hypothesized value of 76 bpm as the population mean.

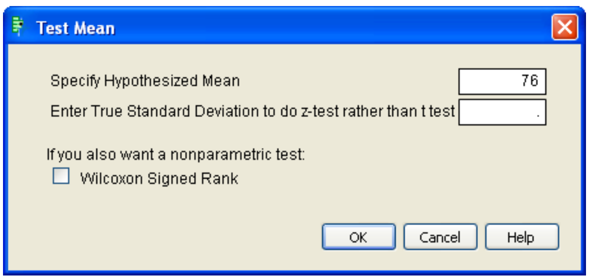
The point of this example is that significance testing and confidence interval construction flow naturally within the software from the initial stages of descriptive analysis. All JMP commands generate graphics alongside numerical results, so that both the descriptive and inferential procedures provide the learner with simple visual imagery to bolster interpretation of outputs. Notice also that the Test Mean dialog box implicitly offers three alternative approaches: tests based on the normal or *t* distribution, as well as the distribution-free Wilcoxon Signed Rank test. This introduces the idea that the ability to infer depends on our data at hand and what we are willing to assume.



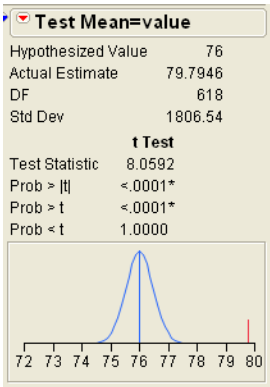
8a: Distribution of Sample Data



8b: Context-Sensitive Menu



8c: Specifying a Significance Test



8d: Test Results

Figure 8: From Description to Inference

Even before formal training in inference, students can be asked to reconcile the 74–78 bpm guideline with this sample of women whose heart rates were nearly 80 bpm in the survey group. How might we account for the difference between the sample mean and the guideline? Would we be willing to attribute the difference to the typical variation among samples? How might we decide?

To assess student progress in this mound, I ask the questions just cited in a lab report assignment. On an exam I ask students to perform a parallel analysis on the data for males 18-25.

3.5 Mound 3: The Logic of Inferential Modeling (4 weeks)

This mound begins with a discussion of (and for some students an introduction to) the scientific method and the logic of inference from sample data. The core concept in this mound is the essential underlying logic of statistical inference. The permutation test is introduced as the canonical technique, first with manipulatives then with software, emphasizing visualization of sampling distributions. We continue with further use of software to simulate random, non-random, and complex samples. Hypothesis generation was previously introduced and now there is more formal coverage of the subject. The course presents inference as consisting of estimation and significance testing—and builds these concepts with resampling and permutation tests, emphasizing the concept of a P-value.

We then cover conventional *t*-tests and interpretation of confidence intervals. As appropriate to the SOS-partners, we can also teach simple modeling using additional theoretical distributions (*e.g.*, Chi-Square, Poisson), and expand communication skills to encompass speaking and writing about inferences results. Especially in SOS courses allied with social sciences, a brief treatment of the concept and use of sampling weights is appropriate here.

JMP includes several native features that facilitate achievement of these particular objectives. Within the interactive platform-specific menus, simulation is straightforward without the need for additional code or external packages. Additionally, there are interactive P-value and power animations linked to *t*-tests. These function much like applets with which readers may be familiar, but make use of the sample data at hand within the context of the analysis being performed. Finally, for the adventurous instructors, it is also possible to write or modify scripts to carry out specific demonstrations.

For example, using the NHANES data referenced in the previous section we investigate causal factors underlying variability in individual health-related measurements (like blood pressure or heart rates). In this P-D-R cycle, students carry more of the planning burden through an assignment in which they develop an analysis plan based on the list of variables available in the dataset. They then select a subsample for model development, and then use JMP's Graph Builder tool to construct a boxplot like the one shown above in Figure 6. In the Graph Builder, users drag and drop candidate factors (gender, race, marital status, etc) to generate side-by-side boxplots. By swapping out different plausible factors, students have the chance to repeatedly think about cause-and-effect relationships and receive instant visual feedback about the models, selecting one for further more formal analysis using the remainder of the NHANES data. Some instructors might elect to apply the sampling weights provided in the dataset (see Figure 3 for an illustration of a JMP dialog accommodating weights).

By this point in the course, students also will be facile enough with the software to undertake modest independent or group-oriented investigations, forming a discipline-specific research question, locating some suitable real data, performing appropriate analysis and reporting the study. Depending on the size of the class, short presentations by student teams would also be advisable both as an assessment tool and as a valuable learning outcome.

3.6 Mound 4: Further Applications of Inference (3 weeks)

The selection of additional inferential techniques should be guided by allied SOS disciplines partners, and thus will depend on the particular setting of the course. Certainly bivariate methods should appear here, and multivariate techniques should also enter the mix. Such topics might include ANOVA, Regression, or another technique commonly applied in the discipline. As in earlier portions of the course we should think about which aspects of each topic are obviated by the software and which ones can recede into the background. The range of illustrative examples is too large to develop specifically here, but examples and topics should be selected according to the extent to which they advance the primary objectives of statistical thinking and practice and accord with the needs of the SOS disciplines.

In this part of the course, one should include substantial student projects that step all the way through the P-D-R process of a good statistical study. This might be a continuation or follow-on to an earlier project or a standalone assignment. A major project serves as an outstanding assessment opportunity. Finally, the course concludes with a more sophisticated treatment of the same important problem with which it began.

4. DISCUSSION

The course sketched here is ambitious and deliberately provocative. Some recommendations are more feasible than others, and some reflect an orientation to social science and business. It is my hope that the framework of constrained optimization is a useful one, and that this proposal is both constructive and in keeping with the spirit of George Cobb's challenge. There is a marked reduction in the coverage of specific methods and techniques, and greater attention to foundational concepts and to real problems that yield to statistical analysis.

Now that a course in statistics is part of so many students' program requirements, the "first" course is no longer an introduction to the field as much as it is the one opportunity we have to make a lasting impact on the way a college-educated generation thinks about interpreting empirical evidence. The introductory course is better conceived of as the *terminal* course for the large majority of students, though clearly some may discover reasons to continue their studies within the field. As such, the content and approach of the course should serve to excite and inspire students. We should choose the elements for this salad with the care and urgency that accompanies a "one chance" context.

A good recipe includes a shopping list and anticipates that cooks will want to modify it according to their tastes and available ingredients. The recipe presented here features variation with populations, variation within and between samples, the logic of statistical inference and discipline-relevant applications of interference. The discussion includes illustrative examples of specific elements that I have chosen, but readers naturally should adapt to their institutions, interests and experience.

Once we have selected ingredients from the freshest and most attractive locally available, they should be cut into lesson-sized slices, and arrayed artfully into mounds. Finally, the salad should be gently tossed in a dressing that blends important and engaging problems, accessible intuitive technology, our enthusiasm for our discipline, and good humor.

5. COLOPHON

This article is based on a paper presented at the Eighth International Conference on Teaching Statistics, Ljubljana, Slovenia in July 2010. The author gratefully acknowledges the invitation to prepare this article and the invaluable guidance of Rob Gould, as well as particularly constructive critiques by two anonymous referees and encouragement by George Cobb, Beth Chance and Allan

Rossman. Robert Carver teaches business statistics to undergraduates and graduate students at Stonehill College and Brandeis University respectively. Questions and comments should be directed to the author at rcarver@stonehill.edu.

6. REFERENCES

Altman, D. G., & Bland, J. M. (1991). Improving Doctors' Understanding of Statistics. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 154*(2), 223-267.

Bartholomew, D. J. (1995). What is Statistics? *Journal of the Royal Statistical Society. Series A (Statistics in Society), 158*(1), 1-20.

Brown, E. N., & Kass, R. E. (2009). What is Statistics? *The American Statistician, 63*(2), 105-110.

Cobb, G. (2007). *The Introductory Statistics Course: A Ptolemaic Curriculum?* Retrieved from http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art1.

Federer, W. T. (1978). Some Remarks on Statistical Education. *The American Statistician, 32*(4), 117-121.

Friedman, J. H., & Stuetzle, W. (2002). John W. Tukey's Work on Interactive Graphics. *The Annals of Statistics, 30*(6), 1629-1639.

Gal, I. (2003). Teaching for Statistical Literacy and Services of Statistics Agencies. *The American Statistician, 57*(2), 80-84.

Hogg, R. V. (1991). Statistical Education: Improvements are Badly Needed. *The American Statistician, 45*, 342-343.

Holmes, P. (2003). 50 Years of Statistics Teaching in English Schools: Some Milestones. *Journal of the Royal Statistical Society. Series D (The Statistician), 52*(4), 439-474.

Kerr, S. T. (1989). Reform in Soviet and American Education: Parallels and Contrasts. *The Phi Delta Kappan, 71*(1), 19-28.

Lesser, L. M. (Ed.). (2007). *Critical Values and Transforming Data: Teaching Statistics with Social Justice* (Vol. 15).

Love, T. E., & Hildebrand, D. K. (2002). Statistics Education and Making Statistics More Effective in Schools of Business Conferences. *The American Statistician, 56*(2), 107-112.

McAlevey, L., Everett, A. M., & Sullivan, C. (2001). Evolution in Business Statistics Curricula: Learning from the 'Making Statistics More Effective in Schools of Business' Conference. *Journal of the Royal Statistical Society. Series D (The Statistician), 50*(3), 321-333.

Meng, X.-L. (2009). Desired and Feared-What Do We Do Now and Over the Next 50 Years. *The American Statistician, 63*(3), 202-210.

Millennium Development Goals Indicators. (2010). from United Nations Statistics Division: http://unstats.un.org/unsd/mdg/Data.aspx

Moore, D. S., Cobb, G. W., Garfield, J., & Meeker, W. Q. (1995). Statistics Education Fin de Siecle. *The American Statistician, 49*(3), 250-260.

Obanya, P. (1995). Case Studies of Curriculum Innovations in Western Africa. *International Review of Education, 41*(5), 315-336.

Peres, C. A., Morettin, P. A., & Narula, S. C. (1985). Educating and Training Undergraduate Applied Statisticians. *Journal of Educational Statistics, 10*(3), 283-292.

Phillips, B. (2001). David Vere-Jones's Influence on Statistical Education. *Journal of Applied Probability, 38A*, 6-19.

Radke-Sharpe, N. (1991). Writing as a Component of Statistics Education. *The American Statistician, 45*(4), 292-293.

Resting Heart Rate. (2010). 2010, from http://www.netfit.co.uk/fitness/test/resting-heart-rate.htm

Rosling, H., Rönnlund, A. R., & Rosling, O. (2004). *New Software Brings Statistics Beyond the Eye*. Paper presented at the Statistics, Knowledge and Policy: OECD World Forum on Key Indicators.

Samsa, G., & Oddone, E. Z. (1994). Integrating Scientific Writing into a Statistics Curriculum: A Course in Statistically Based Scientific Writing. *The American Statistician, 48*(2), 117-119.

Schatzoff, M. (1968). Applications of Time-Shared Computers in a Statistics Curriculum. *Journal of the American Statistical Association, 63*(321), 192-208.

Sharpe, N., De Veaux, R. D., & Velleman, P. F. (2010). Business Statistics. Boston: Addison-Wesley.
Smeeton, N. (1997). Statistical Education in Medicine and Dentistry. *Journal of the Royal Statistical Society. Series D (The Statistician), 46*(4), 521-527.
Tukey, J. W. (1972). *How Computing and Statistics Affect Each Other*. Paper presented at the Babbage Memorial Meeting: Report of Proceedings, London.
Utts, J. (2003). What Educated Citizens Should Know about Statistics and Probability. *The American Statistician, 57*(2), 74-79.
Vere-Jones, D. (1995). The Coming of Age of Statistical Education. *International Statistical Review / Revue Internationale de Statistique, 63*(1), 3-23.
Wild, C. J. (1994). Embracing the "Wider View" of Statistics. *The American Statistician, 48*(2), 163-171.
Yasar, O., & Landau, R. H. (2003). Elements of Computational Science and Engineer Education. *SIAM Review, 45*(4), 787-805.