TEACHING STATISTICAL CONCEPTS, FUNDAMENTALS AND MODELLING

Timothy E. O'Brien
Loyola University Chicago, United States
tobrie1@luc.edu

Over the past several years, statistical educators have been involved in teaching 'service courses,' where our audiences often come from the social sciences and humanities. As such, it is unlikely that these students will ever become 'producers' of statistics – rather, they will be 'consumers' or 'users' of statistics. Thus, the courses we teach these students should reflect this fact – instead of focusing on calculations and derivations, our courses are becoming much more conceptual. This paper highlights some of the author's experiences with the transition from 'number crunching' to 'conceptual' basic statistics courses, focusing on in-class activities, student projects, writing assignments, and using computer packages.

INTRODUCTION

Some time ago, I was helping a Genetics colleague with the analysis of his fruit fly data for which the response variable was binary (success or failure) and the explanatory variable was degrees of latitude. After the data were obtained, the colleague claimed he could "crunch the numbers" himself, and proceeded to produce and analyze his derived 2×2 contingency table in which one row corresponded to "low latitudes" and the other to "high latitudes." He was elated when his analysis produced "statistical significant," the all-important "$P < 5\%$." It was then that I started to appreciate the great disservice we do when we teach "number crunching" statistics courses – courses that emphasize formulas over concepts, rote memorization over reasoning.

After years of teaching introductory statistics courses, and reflecting on the larger issue of statistical literacy in society, I have shifted my focus the past several years away from *whether* to *how*. That is, I have come to realize that every educated citizen does indeed need to have a basic level of statistical literacy, and have started to think of how specifically we can do that. Clearly, when introductory courses are taught emphasizing only the successful use of a given formula or set of steps and failing to underscore the underlying *statistical model*, however, these courses perpetuate the misconception that the key to successfully analyzing a set of data amounts to using *the* "correct formula."

My Genetics colleague no doubt had taken a statistical methods course similar to the one I taught years ago to science and social science graduate students using Ott (1993) and akin to the introductory biostatistics course I currently teach to Biology undergraduate students using Samuels and Witmer (2003). These courses cover the usual topics, including one- and two-sample paired and independent *t*-tests, categorical methods, nonparametric methods, linear regression, and one-way ANOVA. From the student's perspective, these courses do indeed emphasize (through exams and homework) choosing the proper statistical tool from the statistical toolkit and properly using that tool. Of course, these courses also underscore critically assessing any necessary assumptions and conveying conclusions in clear non-technical terms, but the focus is more on methods than on concepts. Also, I do encourage my current Biology students to take at least one follow-up course, and many have gone on to take an Advanced Biostatistics course, in which a great deal of time and effort goes into statistical modelling and the like; the topics covered are given at: http://homepages.math.luc.edu/~tobrien/ABTableofContents.pdf.

In contrast to these Biology students – students who may well go on to perhaps become 'producers' of statistics – I also recently had the opportunity to teach an introductory statistics course to humanities and social science majors ('consumers' of statistics), and it is this latter course and audience which is the focus of this paper.

TOPICS COVERED AND ASSESSMENT

When I first joined the faculty at Loyola University Chicago in 1998, the textbook in use for the Statistical Fundamentals course was Friedman *et al*. (1998), which does an excellent job of downplaying the use of involved formulas and emphasizing underlying statistical concepts. For example, instead of conveying the idea of a sample standard deviation with the formula

$$SD = \sqrt{\frac{\sum (y_k - \bar{y})^2}{n-1}},$$

students are asked to find it using the three-step process:

1. find the sample mean and the deviations for each score,
2. square the deviations, and average these squares, but using (n-1) in the denominator,
3. take the square root of this average.

Students can easily remember this root-mean-square (RMS) process, and this procedure is useful later in the course in the chapter on regression. The book also introduces so-called *box models* as a metaphor for the population under study. For example, in the chapter that discusses Mendel's genetics work with pea plants involving crossing two heterozygous parents (each with one dominant trait, denoted 'A,' and one recessive trait, denoted 'a'), the box model for the phenotype of the next generation contains one 'a' and three 'A's, thereby conveying the probability of 75% of observing a dominant offspring. Box models are also used in Buntinas and Funk (2005) and lay the foundation for statistical modelling; see also Moore and Notz (2006).

Since the text by Friedman *et al*. (1998) is somewhat dated and lacks extensive resources for instructors, and since I wanted to de-emphasize the calculation of standard deviations, test statistics, etc., I used the text by Utts (2005a), which is coupled with excellent detailed resources for instructors and with an activities manual, Utts (2005b). We spent roughly two-thirds of the 50-minute class meetings covering the text (approximately one chapter per class), and the remaining class periods involved in in-class exercises or in the computer lab. Final grades were calculated using the following breakdown.

Midterm Exam	22.5 %
Final Exam	22.5 %
Homework and Quizzes	20.0%
Groupwork and Participation	10.0 %
Mini-Projects	15.0 %
Project	10.0 %

Exams were made up of conceptual questions (short answer, multiple choice and fill-in-the-blank problems) and exercises, and calculators were required for homework and exams (although used much less as for a methods course).

The text by Utts (2005a) can be broken into the following four conceptual parts:

1. Obtaining reliable data (including experiments vs. observational studies, bias, etc.)
2. Representing data graphically (including regression, 2×2 contingency tables, etc.)
3. Basics of probability
4. Estimation (point and confidence intervals) and hypothesis testing.

I felt that the text did a good job of sharpening student's eyes in assessing potential biases in news and other studies and stories (discussed in part 1), paved the way for hypothesis testing (discussed in detail in part 4) by introducing the basics for 2×2 contingency tables in part 2, and did not over-do things with details by just covering the basics of probability. Although the text by Utts (2005a) was the primary textbook for the course, I augmented this material by introducing *box models* (for example to illustrate the Central Limit Theorem for independent Bernoulli trials), by assigning outside readings from Peck *et al*. (2006) and Huff (1993), by using the Minitab statistical package when needed (see the section "Using the Computer" below), and by having students gather and interpret their own data that they found of interest.

CONFUSION OF THE INVERSE, BAYES RULE, AND PROBABILITY TREES

On the first day of class, I circulated a Class Survey and gathered data which was used later in the course. One of the questions of this Survey was the following.

> Suppose 1 in 1,000 people have a disease. A test for it has a 10% false positive rate and a 10% false negative rate. If someone tests positive for the disease, the chances that they actually have it are about: __________%

Approximately 75% of the students in this class answered "90%," and thus were guilty of what Utts calls 'confusion of the inverse' thinking that the question was instead asking, "given that a person has the disease, what are the chances s/he tests positive?" We covered this topic in depth when we covered probability, and defined *specificity* and *sensitivity*, terms that are also useful in the context of drug studies. The manner in which Utts (2005a) introduces Bayes' Rule is by using a 2×2 table; in the above example, the table would be the following.

		How Person Tests		Total
		Positive	Negative	
Person's True Status	Has Disease	90	10	100
	Doesn't	9,990	89,910	99,900
Total		10,080	89,920	100,000

I found it useful to discuss this *table approach* but also to use the *probability tree approach*, wherein the first branch corresponds to the person's true status, and the subsequent branches corresponds to the test results.

Furthermore, this strategy of using probability trees proved useful later in the course in several instances. First, when discussing hypothesis testing and distinguishing between p-values and the probability that the null hypothesis is true, we drew a tree in which the first branch corresponded to the veracity or otherwise of the null hypothesis, and the second branch corresponded to seeing data as extreme as or more extreme that that which we observed. In this manner, students clearly saw that the p-value is not the probability that the null hypothesis is true, a common misconception for beginning students. Probability trees were also helpful for students to visualize the derivation of probabilities for the "Let's Make a Deal" classroom activity as discussed in Utts (2005b, pp. 22-25). Finally, we considered the so-called *random response* strategy given in Samuels and Witmer (2003, pp. 340-341) to ascertain for example an estimate of the percentage of students on campus who are currently using illegal drugs (or some other sensitive subject for which it may be doubtful that selected individuals would answer truthfully when asked). The procedure involves subjects discretely tossing two coins. Subjects who obtained an "H" on the first toss answered the question, "Did you obtain a "H" on your second toss?" Subjects who obtained a "T" on the first toss answered the discrete question (such as "Are you currently using illegal drugs?"). If this latter probability is denoted "p," then students can then show that the probability of receiving a "Yes" response is $(2p+1)/4$, and the probability that a "Yes" responder is in fact an illegal drug user is $2p/(2p+1)$. For example, if the true probability is 1%, then this latter conditional probability is very nearly 2%.

CLASSROOM ACTIVITIES

As an instructor trained in mathematics and statistics, and with only the basics of teaching pedagogy given in my Peace Corps training in West Africa, I found the classroom activities given in Utts (2005b) tremendously beneficial to underscore the ideas and techniques discussed in the classroom lecture and to provide an opportunity for students to learn from one another. For example, students learned about the strategy of staying with one's original choice or switching to another choice in the "Let's Make a Deal" activity mentioned above. On another instance, confidence intervals were illustrated when each student dropped a tack 100 times (actually 5 tacks 20 times) and counted the number of times the prong landed facing up (deemed a "success" for this exercise); this exercise also illustrated how confidence intervals became narrower when the results were pooled over the four students in each group (i.e., when the sample size increase from 100 to 400). In several activity sessions, students read through and discussed abstracts of research articles (or the actual articles themselves) highlighting potential biases, confounding and interacting variables, etc. At the end of these sessions, students would often note how many of these potential problems were ignored by the authors and thus how questionable were the findings given in these studies and reported in the newspaper.

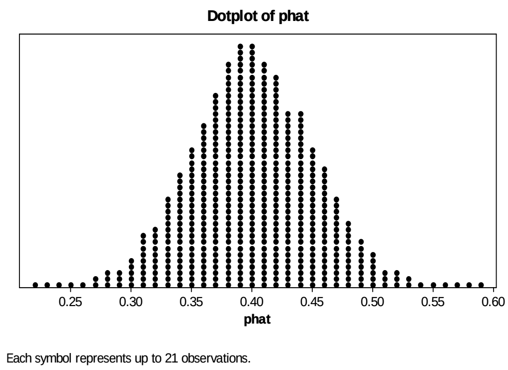
WRITING ASSIGNMENTS AND PROJECTS

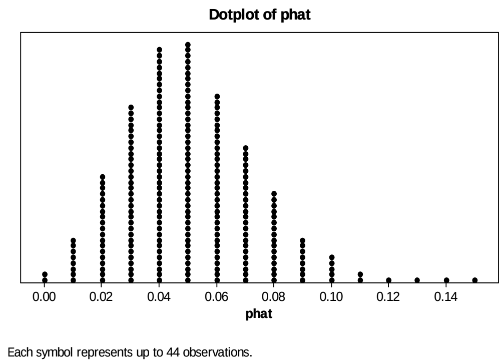
Since my audience was humanities and social science majors, they really didn't mind that 25% of their final grade was based on assignments involving writing, and in that regard this course was writing intensive. The Utts (2005a) text has Mini-Projects at the end of each chapter, and at three instances during the semester, I asked students to complete one of these for a subset of the chapters. For example, one of these asked students to find a research article discussing an observational study, and to write a 1-2 page summary of the study and to mention potential biases and the possible ramifications of these. As an extra-credit assignment, I gave students the option of completing the second assignment given in Jordan (2004) in which students are asked to write a letter to their father explaining and interpreting the p-value associated with a randomized placebo-control study in layman's terms and to discussing its importance. The final course project/paper required that students obtain their own data on a quantitative variable for each of two groups, summarize the data and testing hypotheses, and write up their findings.

USING THE COMPUTER

Numerous concepts are best illustrated using an easy-to-use statistical package such as Minitab. For example, using the following Minitab commands, my students easily produced the histograms below, thereby illustrating the Central Limit Theorem for a Bernoulli proportion when n = 100 and p = 0.40 (left histogram) and when p = 0.05 (right histogram).

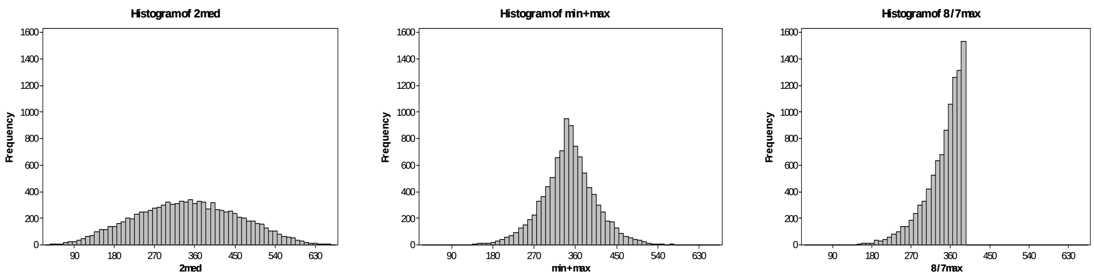
Calc -> Random Data -> Bernoulli -> 10000 rows -> store in columns C1 – C100 -> probability of success 0.40 -> OK. Next, label column C101 phat, then Calc -> Row statistics -> click mean -> input variables C1-C100 -> store result in phat (or C101).





Even though the latter situation (n = 100, p = 0.05) barely satisfies the usual "largeness" (of sample size) criterion (that is, both $n*p$ and $n*(1-p)$ at least 5), students easily see that the normal approximation is becoming dubious for this case.

Another illustration corresponds to the "German tank exercise" discussed in Bullard (2003) in which sets of 7 randomly chosen integers (used as a proxy for serial numbers found on the captured German tanks) between 1 and N, and where the goal is to obtain a "good" estimator of N; I chose N = 344 for the in-class exercise. My students focused on the estimators $T_1 = \bar{x}$ (the sample mean), T_2 = max, T_3 = median, $T_4 = \bar{x} + 2SD$, $T_5 = \bar{x} + 3SD$, $T_6 = 2*\bar{x}$, T_7 = 2*median, T_8 = min + max, and T_9 = 8/7*max, and the class conversation ultimately centered on unbiasedness and small variance. Then, in a subsequent computer lab, we produced the following histograms (for T_7, T_8, and T_9, respectively) and summary statistics (for $T_1 – T_9$).



Descriptive Statistics

Variable	N	Mean	StDev	Median
mean	10000	172.53	37.78	172.57
max	10000	301.24	38.37	312.00
median	10000	172.53	57.65	172.00
xbar2SD	10000	366.25	56.93	372.10
xbar3SD	10000	463.10	74.08	469.04
2xbar	10000	345.06	75.55	345.14
2median	10000	345.06	115.31	344.00
min+max	10000	345.07	58.07	345.00
8/7max	10000	344.27	43.86	356.57

After viewing these results, students can easily appreciate that T_9 = 8/7*max is "best" in terms of smallest variance amongst the unbiased choices. But we can go one step further.

In the context of the above Bernoulli examples with p = 0.40 and p = 0.05, we can also construct the approximate 95% confidence intervals using the Wald method ($\hat{p}$ ± 2SE) and count the number of times these intervals contain the true p; this is achieved using the following.

Next, label C102 SE, and follow Calc -> calculator -> store result in C102 -> expression sqrt(C101*(1-C101)/100). Next, label column C103 L, column C104 R, and column C105 covers40, and follow Calc -> calculator -> store result in C103 -> expression C101-2*C102
Calc -> calculator -> store result in C104 -> expression C101+2*C102
Calc -> calculator -> store result in C105 -> expression ((C103 LE 0.40) and (0.40 LE C104))
Then, Calc -> column statistics -> mean -> input variable C105 gives the following
`Mean of covers40 = 0.9564`, meaning that 95.64% of my intervals contained the true 40%.

For the case of p = 0.40, 95.64% of our intervals covered the true value, and this calculation helps students to understand the correct (Frequentist) interpretation of confidence intervals. This Wald procedure is called into suspicion for the case where p = 0.05 since in that case the actual coverage rate of the 95% confidence intervals was only 87.27% and since 25.72% of these intervals also contain zero (so that the interval includes negative estimates for p). The class also discussed the Score confidence intervals method discussed in Santner (1998), and we found these intervals superior since none contained zero and since the nominal coverage rate was 96.44%.

Finally, we returned to the German tank exercise to evaluate the performance of the estimators T_6 – T_9 in terms of coverage probabilities of the corresponding 95% Wald-type confidence intervals. Based on 100,000 replications, these percentages were 95.12%, 95.95%, 93.74% and 94.78%, respectively. That each of these values are nearly all the same and nearly equal to the nominal 95% indicates that none is preferred to the other based solely on coverage of the respective confidence interval. This helps students understand that some uncertainty and controversy exists in terms of evaluating estimators.

CONCLUSION

Even though there is some truth in the statement that "all [statistical] models are wrong but some are useful" (Box, 1979), a course in statistical methods which ignores statistical modelling and which fails to underscore the underlying concepts is easily seen as a jumble of formulas and usually leaves students with the impression that statistics is merely "number crunching." On the contrary, the real strength of statistical science is its ability to develop new methods to directly aid researchers and decision-makers, and we need to emphasize this in our courses.

By way of an epilogue, encouraged by my successes in this course and some of my students, I have agreed to teach another course on statistical concepts next semester, this one on *Statistics and Medical Ethics*, and using the texts of Angell (2004), Avorn (2004) and Crossen (1996). Should be fun.

REFERENCES

Agresti, A. and Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119-126.

Angell, M. (2004). *The Truth About the Drug Companies*. New York: Random House.

Avorn, J. (2004). *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. New York: Alfred A. Knopf.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer, and G. N. Wilkinson, (Eds.), *Robustness in Statistics*. New York: Academic Press.

Bullard, F. (2003). "The German Tank Problem," presented at the Beyond the Formula Conference, Rochester, NY, http://web.monroecc.edu/manila/webfiles/beyond/2003S022S071Bullard.pdf

Buntinas, M. and Funk, G.M. (2005). *Statistics for the Sciences*. Belmont, CA: Thomson.

Crossen, C. (1996). *Tainted Truth: The Manipulation of Fact in America*. New York: Simon and Schuster.

Freedman, D., Pisani, R. and Purves, R. (1998). *Statistics* (3rd Edition). New York: Norton.

Garfield, J. (2004). Becoming an effective teacher of statistics. *Stats Magazine*, 40, 8–11.

Huff, D. (1993). *How to Lie with Statistics*. New York: W. W. Norton and Company.

Jordan, J. (2004). Tales from the Frontline: A Pilot Study of Writing Assignments. Handout at Innovative Ideas for the Statistics Classroom Session at the Joint Statistical Meetings, Toronto, CA.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 65(2), 123-165.

Moore, D. S. and Notz, W. I. (2006). *Statistics: Concepts and Controversies* (6th Edition). New York: W.H. Freeman and Co.

O'Brien, T. E. (2002). Hypothesis tests, confidence intervals and common sense. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Ott, R. L. (1993). *An Introduction to Statistical Methods and Data Analysis* (4th Edition). Pacific Grove, CA: Duxbury.

Peck, R. Casella, G., Cobb, G., Hoerl, R., Nolan, D., Starbuck, R. and Stern, H. (Eds.) (2006). *Statistics: A Guide to the Unknown* (4th Edition). Belmont, CA: Thomson.

Rossman, A. J. and Chance, B. L. (2005). "Teaching Statistical Concepts with Activities, Data and Technology." Short course handout at MathFest, Albuquerque.

Samuels, M. L. and Witmer, J. A. (2003). *Statistics for the Life Sciences* (3rd Edition). Upper Saddle River, NJ: Pearson Education (Prentice-Hall).

Santner, T. J. (1998). Teaching large-sample binomial confidence intervals. *Teaching Statistics*, 20(1), 20-23.

Utts, J. M. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74–79.

Utts, J. M. (2005a). *Seeing Through Statistics* (3rd Edition). Belmont, CA: Thomson.

Utts, J. M. (2005b). *Activities Manual for Seeing Through Statistics* (3rd Edition). Belmont, CA: Thomson.