Some Embarrassing Questions for Teachers of Statistics

Larry Weldon
Department of Mathematics and Statistics
Simon Fraser University
Burnaby, B.C.
Canada. V5A 1S6

Embarrassing Questions?

Generally, I think teachers of statistics should be embarrassed if the dogma they are teaching is patently nonsensical, clearly out-of-date, or of no practical use. They should also be embarrassed if the statistical dogma that would be most useful to students is omitted from the statistics courses. I have arranged my more detailed comments according to these categories of atrocities.

1. Patent Nonsense:

1.1 Is "variance" a reasonable measure of variation?

No! The units are inappropriate. SD has the right units. The nomenclature is unfortunate since it would suggest that the answer is Yes. The reason this happened is that in pre-computer days, it was important to have things that could be easily computed by hand, and variances were easier than SDs because they avoided the square root operation. This is no longer a convenience of any importance.

Mathematicians may argue that variance has many convenient properties that the SD does not have, such as its ANOVA partition; but this argument is about how theorems are proved, and not about the definition of appropriate summary measures (i.e. statistics).

1.2 What aspects of an analysis-of-variance table convey useful information?

The P-value. Everything else is for the hand-calculation routine for computing the P-value. But computers now do the calculation, and even statisticians no longer compute this P-value by hand. The logic of analysis of variance is useful for both scientists and statisticians to understand, but the analysis of variance table does not help explain this logic - in fact, it obscures it!

(The key, of course, is that the standard error of a mean is the SD of the individual observations divided by sqrt(n), and this fact is what leads us to use the size of a variance ratio as an indicator of the evidence against the equality of group means.)

The argument for banishing most of the analysis-of-variance table from statistical reports of data analysis, except the P-value, should not be taken as a criticism of the testing procedure itself. It is simply that the demonstration of the calculation procedure in particular cases is of no interest to most audiences, including statisticians and scientists..

1.3 Is the P-value a reasonable quantity to compute in evaluating the credibility of an hypothesis? Why is .05 considered an appropriate critical P-value?

It may be argued that it is reasonable in the sense that a decision on credibility clearly should be based on some monotonic function of the P-value. The more abnormal the data under the hypothesis, the stronger the evidence against the hypothesis, and the P-value does move in the right direction. However, so do many other monotonic functions of the P-value - so this argument alone is not a compelling one in favor of the P-value. It is sometimes argued that the P-value measures the probability of the sample under the hypothesis, but actually it measures the probability of the observed sample and many others not observed. Some argue the likelihood ratio is more reasonable, and I agree. However the tradition of P-value will be with us for some time yet, and there is no denying that it has pragmatic usefulness, so it is good to teach it anyway. We should be aware though that the world is waiting for a better method, and we should not be too dogmatic about it with students.

One aspect of this dogmatism that really should be noted for students is the arbitrariness of the conventional .05, .01, and .001 critical P-values. The choice among these or other critical P-values is arbitrary, (even if guided by a subjective opinion about costs of inference errors), and so it does not make much sense to worry about whether we have the observed P-value computed correctly to the third decimal place, or even to the second decimal place. To emphasize this point, should a statistician who goes through life using alpha=.001 be paid more than one who uses alpha = .10, because he/she makes fewer type I errors? No, because he/she will make more Type II errors, and we don't have any routine way of choosing the best mix of these, since different situations call for different mixes. Even in a particular instance, there is not a routine way of choosing an optimal mix, since the choice of loss function is not routine.

We must be candid about the ad-hoc nature of traditional hypothesis testing, and not portray it as an optimal decision-making procedure, because it has many serious failings as such.

1.4 What is "theory" as opposed to "practice", and how does this distinction apply to the discipline of statistics? Do we teach this theory to anyone? Should we teach it to everyone?

In any discipline, the "theory" is the collection of principles that are used again and again in different situations to solve the problems of interest to the discipline. I use the phrase "generally applicable concepts" to describe these "gems" that form the basis of a discipline. In other words, theory is not really "opposed" to practice at all, but rather the collection of things that are most "able-to-be-practiced" in contexts that will arise in the future. If we teach "practice" without teaching theory, we are only teaching how to operate in known contexts of the past. This can be useful in many situations, but a university education usually emphasizes the ability to adapt to the unknown, rather than only to take advantage of what is already known.

In Statistics, the theory consists of the "big ideas" that help us time and again to sort out the complexities involved in interpreting data that is subject to unexplained variation, or "randomness" as we like to label it. Some of these ideas are best described mathematically, but many have little to do with mathematics. For example, the idea that a model simplifies the reality it describes, is an idea that is familiar to architects, political scientists, and dentists, and they do not need to be mathematicians to understand this idea. The point is that to focus on statistical theory as a subset of mathematics is guaranteed to convey a wrong impression of what is important to statistical theory. We have to teach students that statistical theory is not mathematics, but rather statistics, a different discipline. Applied scientists, engineers, and investment analysts each make serious use of mathematics without considering that the theory of their disciplines is primarily mathematics -- what self-respecting communications engineer would agree that he should be examined in his discipline by a mathematician? The point here is that mathematics is a powerful tool useful to many disciplines, including statistics, but the "theory" of these disciplines is not to be judged by the quality of the mathematics used, and neither should statistics.

We need to abolish the concept of "service" courses, in favor of "basic" courses, and to teach basic statistical theory to everyone who wants to know something about the discipline of statistics, including the mathematicians.

Anyone who wishes to become a career statistician should take additional courses in mathematics and statistics. But career statisticians still need to understand the non-mathematical basics of the discipline, and their introductory courses must not skimp on these aspects in favor of the mathematical devices. The big picture must be in view before the mathematical details make sense.

The proposal here is that all students in statistics courses must be taught the basic theory, and that this need not involve mathematics beyond the high-school level. Students seeking an in-depth treatment must take additional courses in statistics, and some of these additional courses will require a stronger mathematical background. But it is important that the "in-depth" treatment begin with the big ideas of statistics, and most of these can be conveyed without higher mathematics. Using calculus in the introductory course in statistics obscures the central themes of statistics.

1.5 What is the purpose of a confidence interval? Does it do the job that is needed?

A confidence interval provides an interval estimate of a parameter value. However, contrary to some statistical theorists, most users of statistics really just want to know where the values of the distribution tend to lie, and they are not so interested in the parameter of the distribution. A confidence interval has a role to play but it is a fairly minor one. Most students think that a confidence interval does tell them where the distribution is, and not just its mean, and it is not surprising that this misconception prevails. Students apparently perceive that an estimate of the whole distribution is a natural thing to report from a sample, and one can easily accept this perception. If we only teach confidence intervals and ignore the estimation of whole distributions, we are bound to mislead students.

The focus of theoretical statisticians on parameter values derives from a desire to describe, in mathematically tractable terms, the complexity of frequency distributions. As our ability with computers to cope with whole distributions, instead of only their parametric approximation, our need for parametric inference will decline. Then confidence intervals will fade even further in importance, at least by those users who really understand what they describe.

1.6 What is the purpose of a hypothesis test? Does it do the job that is needed?

It is sometimes described as a decision-making technique. But it does not really make decisions, it only weighs evidence against the null hypothesis. We must admit to students that "accepting a null hypothesis" is very different from proving, or even deciding, that the hypothesis is true, or that a hypothesis is believable. Similarly, we must point out that "rejecting the null" can be a "nit-picking" affair, in the sense that the null may be, for all practical purposes, true, and still be rejected, if the sample is large enough. Without a serious discussion of loss functions (or utility functions), a proper decision-making procedure cannot be devised. We have either to refrain from describing hypothesis testing as a decision-making procedure, or else go into the details. Of course, the role that hypothesis testing does properly play is of value to scientists and others, but it is not really a decision-making procedure, as it is so often portrayed.

1.7 Is regression a curve-fitting technique?

Only in a very specialized sense, as a prediction curve. For a given pair of variables there are 2 ways to regress one of them on the other, and the choice produces two very different "fits". So *the* regression fit to a bivariate data set is not a well-defined concept. Regression is better described as a prediction technique, or as a method to determine predictive relationships.

If there is need for a fit to bivariate data, but where no prediction is contemplated, then the fit should be assessed using distances of the data points from the curve that are perpendicular to the fitted curve.

1.8 If there are two ways to measure one thing, is the correlation between the two measures a reasonable guide to the degree to which one measure can substitute for another?

No. The correlation is independent of location and scale, and location and scale are important ways that two measurement techniques can be dissimilar. Some measure based on the difference in the two measurements is needed, such as the so-called “technical error of measurement”.

1.9 What fact do we need to know in interpreting the magnitude of the correlation between two variables?

We need to know how the range of the variables have been determined. Has a population been sampled (usually not)? If not, how have the cases under study been selected. If we want to know the correlation between height and weight, it matters a lot whether we include children, both sexes, athletes, etc. In fact we can manipulate the correlation by choosing the population sampled. (Note that this same problem affects the meaning of R-squared as a measure of the extent to which X determines Y in a regression.)

2. Out-of-Date:

2.1 Are probability models needed for most data analyses?

No! The need for anything other than the normal distribution is rare, at least in most traditional data analyses. Nonnormality (often perceived as skewness) is usually dealt with by symmetrizing transformations, and treating the data as if it were normal. Poisson and Binomial probabilities are often treated as approximately normal. Testing for normality is rarely more rigorous than checking the linearity of a normal probability plot.

In other words, most data analyses require little more probability than the normal distribution. Of course, students need to understand independence and mutual exclusiveness, but these ideas do not require much sophistication about probability models. Basic probability ideas like "long run relative frequency" and "sampling variability" need to be understood well, but these ideas can be thoroughly explained without any reference to negative binomial or gamma distributions.

The argument here is not that probability modeling has no use in data analysis, but rather that the usual procedures do not require an understanding of these models for their execution. We argue in 4.10 below that more use could be made of probability modeling, for statistical analysis.

2.2 Is it necessary for students to be able to compute simple statistics -- like standard deviations and F-statistics -- quickly and accurately?

Only in the sense that they have to know how to use statistical software. It is not useful to have students using calculators to demonstrate that they know how to use a formula or procedure. Computers do this. Students have to know when certain calculations are appropriate, what outcomes are reasonable in view of the data, and how to interpret results. But the mechanics of the procedure is something they will never have to reproduce without use of a computer.

2.3 Do students know how to make use of residual plots?

Modern texts do emphasize that residual plots are important in evaluating a model, at least in the regression setting. This has come about since computers have made the construction of residual plots very easy to do, and feasible to do routinely. But how many students understand how to use the residual plot to improve the model? The iterative nature of model-fitting, emphasized by George Box, does not receive much attention in "theory" courses. Yet iterative modeling it is an important part of the generally applicable concepts of statistics, and should be included in the basic courses. This is just one aspect of the important impact of computer developments on statistical theory. This idea is repeated as a small part of the question 2.4.

2.4 Has the computer revolution had any effect on the discipline of statistics, as taught in university courses? Are formulas for hand-calculation of sums of squares still taught? What use is made of the following computer-intensive methods in elementary statistics courses?

i) computer graphics for data summary
ii) computer graphics for data analysis
iii) computer simulations
iv) resampling methods such as the bootstrap
v) iterative approach to modeling and analysis
vi) sorting data by variables of interest

The effect on university courses has been slight, judging by the texts that are commonly used. While "service course" text books have become a much larger proportion of the statistics textbook market, as a result of the demand by non-statistics departments for application-oriented courses in statistics, it is not true that these textbooks have changed much in their list of topics: histograms, sampling distributions, confidence intervals, hypothesis tests, regression and anova. But these courses **should** be responding to the changing environment of the computer age. We don't need hand-calculation formulas for sums of squares that obscure the nature of the thing calculated. Moreover, the items listed in i) to vi) above are at least as important as the traditional topics and should be included in a first course in statistics.

2.5 Is it preferable to use a parametric model, rather than a nonparametric model, when we are summarizing a predictive relationship?

Not usually. Most models are for empirical description, rather than mechanistic explanations. When we fit a quadratic curve to some sales trends, we do not mean to imply that the nature of sales is inherently quadratic, but merely that this curve provides a close fit. Techniques such as "loess" (Cleveland, 1993) are general purpose smoothers that perform the same function without requiring any assumption about the parametric form of the relationship under study. The need for parametric models in pre-computer days was motivated by the difficulty of communicating relationships graphically -- but with computers, the loess kind of fit can be easily presented and can also be easily comprehended. These nonparametric fits need to replace some of the time spend on teaching parametric fits.

3. Of No Practical Use:

3.1 Is it important to teach students how to test that a statistical model is a correct one?

All models are wrong! The utility of a model is in the degree to which it simplifies the reality it represents, not in the extent to which it duplicates this reality. A perfect model is not really a model. (Even a model airplane is too small!) When we encourage students to check assumptions, do we really expect them to ensure that the assumptions are exactly valid? Of course not. What we hope is that the violations of the assumptions are such that they suggest an improvement to the model, a better fit without adding too much complexity. To test "goodness-of-fit" does not make as much sense as to measure the goodness-of-fit, but there are no standard techniques to do the latter.

3.2 Do students learn the BIG ideas of statistical theory?

Apparently not. How many students realize that a graphical summary of a data analysis is usually the most effective, or that the response rate in a sample survey does not tell us much about the danger of response bias, or that a confidence interval does **not** tell where the distribution roams, or that a hypothesis test cannot prove that the null hypothesis is true. Most students think the big ideas of statistics are the formula for the normal density and that $P<.05$ proves the null is false. This bias toward the importance of mathematical things is hazardous for instruction in statistics per se.

The first course in statistics should make clear those statistical concepts that are most useful for modern practice. Non-mathematical students will expect this. The students who are strong in mathematics will never realize what the big ideas are, if the course requirements do not pull them away from their dependence on algebraic formalisms.

3.3 Is instruction in statistical computing specific to one brand of software, or can it be used with any statistical software?

University courses in statistics should not put too much emphasis on a particular computer package. Once students know one package, courses should allow students to use any package they have access to. This approach emphasizes that statistics courses should teach the big ideas, not the details of particular software.

3.4 Are students' minds actively engaged while attending statistics lectures?

From what I hear from students, usually not! The textbooks, and some instructors, seem to convey the message that statistics exams test whether students have memorized the standard procedures of statistical inference. So students are filtering out comments about concepts and judgment, since they assume this is not examination fare. This does not encourage two-way communication in the classroom, or even successful one-way communication of ideas.

3.5 Are teachers of statistics experts in application of the theory?

Many have gone from Ph.D. in statistics to teach statistics without ever having to apply what they have learned. It is likely that they will not know what the "generally applicable" concepts are, since they have never applied any concepts. One remedy for this is for the instructor to get involved with data-based projects done by the students -- everyone can learn about statistical theory this way, including the instructor!

3.6 Do students recognize the great value to their careers of a thorough understanding of statistics?

Service courses in statistics are usually compulsory, so we cannot take much encouragement from the large numbers in these courses. Optional selection of service courses is not nearly so common.

Statistics as a major attracts a fairly small proportion of students, especially when compared with huge numbers taking all those service courses in statistics. Many students have supposedly been familiarized with the discipline, but very few decide to choose it as a career. One wonders what does motivate those few who do choose statistics? Is it the prospect of gaining an employable skill? Is it a way to use a training in mathematics without having the talent for abstraction that pure mathematics requires? Is it a love of arcane formulas and an inclination for compulsive checking of arithmetic? If so, these motivations should concern us. One hopes that students are drawn by the generality of the concepts for the conduct of good research, and for the opportunity that expert statisticians have, to get involved in collaborative research with others. Our courses should appeal to these latter motivations.

Statistics has the notoriety of being loved by believers and hated by everyone else. But the subject is not really so crisply defined as to make this a reasonable state-of-affairs. Apparently, we are teaching statistics as if it were a religion, instead of presenting it as a professional tool. The theory of statistics is still very crude and in desperate need of fresh ideas. Teachers who present statistical theory as a tidy collection of optimal strategies are not doing the discipline a favor -- they are misleading the potential users and encouraging skepticism about the real value of the subject. There are good reasons why so many researchers groan at the thought that they will have to use some statistical tools -- teachers of statistics have to understand what these are and try to make suitable adjustments to their teachings.

3.7 Do students receive adequate training in integration and differentiation so that they can solve real-world statistical problems creatively?

One course in calculus is probably enough for 99.9 percent of the problems that statistical practitioners will face. Programs like MAPLE that do mathematical manipulation, differentiation and integration symbolically can bypass some of the training that used to be thought essential, like substitution tricks for integration, and memorizing dozens of formulas. "Resampling statistics" techniques (simulation, bootstrap) allow avoidance of calculus in the solution of many problems. It is rare these days for a statistical practitioner to need a detailed knowledge of calculus in his/her work.

3.8 Is it important to know how to use t-strategies to allow for variability in the estimation of standard deviations?

Not really. The arbitrariness of the critical P-value suggests that computation of exact P-values is not necessary, as long as approximate ones are available. In the case when the standard deviation of the population sampled is unknown, and the sample is too small to provide a reasonable estimate of the standard deviation, we are trained to use t-strategies. But the normality of the population can never be checked with a small sample, and while it these t-strategies are usually robust to non-normality, it is obvious that just a little contamination by genuine "outliers" can upset the apparent precision of the data, and subvert the t-strategy. The clever devise invented by Gossett has its uses but it is not as generally useful as textbooks would suggest, especially given the additional complexity that it introduces to both instruction and practice.

4. Glaring Omissions

4.1 What are the most widely used techniques in real-world data analysis?

If one judges by the use of statistics in the media, graphical displays of time series, sample survey summaries, and bar charts would be most widely used. If one considers industrial uses of statistics, one would have to add control charts. For retail outlets of all kinds, a record of the daily sales figures is almost always examined, and this is also a kind of time series.

When we teach basic statistics, we never talk about time series; we say very little about how to summarize survey questionnaires (e.g. how does one summarize a 10 question survey that has a five-scale response: DS, D, N, A, AS?); graphical summaries, if they are discussed at all, are limited to histograms, stem-and-leaf plots, and scatter diagrams; and control charts are omitted entirely. Why do we leave out the things that would be most useful?

4.2 Which is the best way to convey a scientific result:
i) a parametric summary
ii) a verbal summary
iii) a graph ?
How do we examine students' ability to do this?

Most textbooks emphasize i), but ii) and iii) are clearly better and more important. A statistical analysis that stops at i) will not influence anyone.

Exam questions usually require a calculation of some sort. But how many ask to explain a given result in words, or to suggest appropriate graphical displays?

4.3 What is the first step in a data-based project? What is the second step? Do we teach these steps?

Step 1: Decide on the objective of the project, and write it down.
Step 2: Collect the data
Step 3: Check the data for anomalies

We don't usually teach these things in statistics courses. Why not? There are in fact things to be taught about each of these steps.

4.4 Do students know what to do when comparisons are confused by a known confounding factor?

Simpson's paradox is a useful teaching tool, because it shows how extremely misleading an unaccounted-for confounding factor can be. Of course, the point of this is not the paradox itself, but the fact that observational studies have special problems for the interpretation of comparisons, and that when data is available for suspected confounding factors, one must adjust the comparisons by removing statistically the influence of the confounding factor. In the case of categorical variables (the context in which Simpson's paradox is usually portrayed) this means examining comparisons within each value of the confounding factor. How often do we convey this important point in the context of comparisons from observational studies?

4.5 Do students know how to decide if a survey response percentage is high enough for valid inferences based on the sample alone?

This is a matter of judgment usually. We must guess the extent to which the propensity to respond will be related to the response itself, for this is the thing that causes non-response bias. Consultees who ask if 50 percent response is large enough to give reliable responses, are asking the wrong question. Response bias can be serious with 90 percent response if the respondents are different enough from the nonrespondents. (For example, consider the question: Have you ever committed a crime that, if you were caught, would have resulted in a jail sentence of at least one year?)

One other strategy that is sometimes relevant here is to use whatever information is available abut the non-respondents (address?) to compare the respondents and non-respondents.

Are students taught these important ideas? Why not?

4.6 Is it necessary for teachers of statistical theory to know how to apply the theory?

YES! All students of statistical theory have to know how to apply the theory, even those who aspire to teach statistics in a university setting. Teachers should be helping students to relate the theory to the practice of statistics and they cannot do this unless the teachers have actually done some application. So much of the theory of statistics is only understood through the medium of applications that the full understanding of the theory from passive classroom experience is virtually impossible. How can a student experience the reality that "all models are wrong" without having to choose a model to describe a particular phenomenon of interest?

4.7 How should a first year service course in statistics differ from a first year course in statistics for statistics majors?

Not at all! Statistics majors should take more statistics, not different statistics. The first survey-style course in statistics should outline the field and hit the big ideas. Most students at this stage are undecided about their specialty in any case, and we should assume that anyone is a candidate for a career in statistics. Even if only a few do choose this option, by giving a serious, but broad introduction to statistics, all will be given a proper perspective about the subject, and there will be more people who can make good use of statistical experts.

4.8 Do students of statistics know how to explain the rationale of statistical methods to non-statisticians?   Do they need to?

As was mentioned in an earlier section, a result from a statistical analysis that cannot be communicated in words or pictures will have little impact. Moreover, if we force students to verbalize their understanding of statistics, we will be helping them to relate the symbolic formulas to their general intelligence, instead of partitioning their brains into "stats" and "everything else".

4.9 What statistical method has had the greatest impact on world economies? Do we teach it in our basic course?

Reasonable candidates would be time series forecasting, quality control charts, and puplic opinion polls. We should include these topics in our most elementary courses.

4.10 Is data always numeric?

Statisticians tend to think that data must always be numeric, since all our formulas assume this is the case. But we train many statisticians with probability models, even though we do not discuss how to relate these models to statistics problems. One way to do this is to think of the observation of a phenomenon, like an epidemic outbreak, an traffic bottleneck, or market fluctuations, as "data", and then our job as statisticians is to provide a simple explanation for the phenomenon. This would normally be done either algebraically or with a simulation if the system modeled is complex. The charming field of applied probability modeling might be more popular if its utility as a statistical tool were more widely understood.

5. Corrective Measures:

5.1 Is there anything that a teacher can do to improve the impact of statistics courses when all the textbooks are so old-fashioned?

Use the textbooks for reference, but present statistical techniques as they arise naturally in a project-based course. Projects can be discussed in class before, during and after the time these same reports are submitted for marking.

5.2 How do students learn to approach problems when most of their training is technique-based? How do students learn about the primary importance of data quality for interpretation of data? How do students learn to relate their verbal intelligence to their numeric and algebraic intelligence?

See 5.1 above

5.3 On what basis does a student express his or her pride in a familiarity with the discipline of statistics?

With guided experience with data-based projects, a student can learn about the pitfalls of a naive approach, and also how to conduct data-based studies efficiently. Since this knowledge is applicable to many different fields, the student becomes broadly educated and is able to appreciate and communicate the key role of data-based research. It can then be a source of pride since it would be recognized as important to almost everyone, like literacy and interpersonal skills.

References:

1. Cleveland, William S. (1993) Visualizing Data. Hobart Press, P.O. Box 1473, Summit NJ USA 07902-8473.

Contact information:

K.L. Weldon
Department of Mathematics and Statistics
Simon Fraser University
Burnaby, B.C.
Canada V5A 1S6

Phone:	604-291-3667	(B)
Phone:	604-943-7962	(R)
FAX	604-291-4947	(B)
FAX	604-948-2685	(R)

e-mail: weldon@sfu.ca

(KLW 95/08/05)