LETTERS TO THE EDITOR

Letters to the editor will be confined to discussions of papers that have appeared in *The American Statistican* and to important issues facing the statistical community. Letters discussing papers in *The American Statistician* must be received within two months of publication of the paper; the author of the paper will then be given an oppurtunity to reply, and the letters and reply will be published together. All letters to the editor will be refereed. Corrections of errors that have been noted in papers published in *The American Statistican* will be listed as corrections at the end of this section.

Thanks for the great article "Karl Pearson and R. A. Fisher on Statistical Tests, A 1935 Exchange from Nature" by Henry F. Inman [February 1994, pp. 2–11]. This article has helped to fill several gaps in my understanding on the controversy on the tests of hypotheses. I am truly grateful for the research and effort made by the author to provide a greater insight into this topic. While pursuing a further understanding on this topic a couple of years ago, I had several discussions with Dr. W. Edwards Deming. An important point to emphasize regarding this debate is a point made by Dr. Deming in a letter to me dated August 26, 1991. Dr. Deming stated:

> I can tell you, though, that I studied with both of them [J. Neyman and E. S. Pearson] in London in 1936. I can assure you that their test criteria had nothing to do with statistical inference. Inference means prediction. They never once thought about uses of data for prediction. The same statement could be made on most statistical writing today.

Dr. Deming is referring to what he has called the enumerative study versus the analytical study. His caution should be remembered by all of us applying statistical data analysis.

Eileen J. Beachell
Quality Disciplines
West Bloomfield, MI 48323

MOORE, DAVID S. (1993), "THE PLACE OF VIDEO IN NEW STYLES OF TEACHING AND LEARNING STATISTICS," *THE AMERICAN STATISTICIAN*, 47, 172–176

I would like to add to the instructional strategies suggested in this paper when segments from the *Against All Odds* videos are used as part of lectures to large groups of students. In Australia, we typically give two-hour lectures on basic statistics to groups of 200 or more students. It is not possible to conduct any kind of discussion in this environment because the students cannot hear what other students are saying, when they attempt to speak either to the teacher or to the whole class.

In this situation, showing segments of the videos is an excellent way to break the monotony of a two-hour lecture during which there are very limited opportunities for interaction. Nevertheless, the students do want to discuss the videos, so I have devised the following procedure.

Before the video is shown, the students are given a sheet of questions that they are asked to read through. This helps them to focus on important aspects of the segment. Some of the questions are factual, others are designed to generate discussion. Selected students are also given a transparency with one or two question numbers on it. As the video segment is being shown, students can raise their hand to indicate that they would like a pause. At these times and at the end of the segment, students discuss the questions with their immediate neighbors and write answers on the question sheet. During this time I take the opportunity to talk to some groups of students who are accessible. In due course the students with transparencies bring them to the overhead projector and show their responses. Sometimes groups of students are asked to hand in their responses as part of homework assignments.

This may sound rather contrived, but once they become accustomed to the routine, students enjoy the opportunity to make a contribution to proceedings.

I would like to add one other point. Increasing numbers of our students are not native English speakers. These students have great difficulty in understanding the videos in the classroom situation described above. (They probably have great difficulty understanding me also.) However, many of them clearly view the videos (many times) after class and eventually produce excellent answers to the questions. Thus in addition to learning about statistics from the videos, they improve their English language skills.

Ruth Hubbard
School of Mathematics
Queensland University of Technology
Brisbane, Australia

COMMENT AND REPLY TO CHRISTENSEN, R., AND UTTS, J. (1992), "BAYESIAN RESOLUTION OF THE EXCHANGE PARADOX," *THE AMERICAN STATISTICIAN*, 47, p. 311: COMMENT BY ROSS AND REPLY

The claim that you should just accept the first value is incorrect. Suppose that you are handed two envelopes, each containing a check. Let A and B be the unknown values of the checks (although the paradox assumes that $B = 2A$, this is not essential in finding a strategy that beats accepting the first value seen), and assume that $A \neq B$. You are to randomly choose one of the envelopes, open it to see the amount of the enclosed check, and then decide whether to accept it or to take the other check.

It is not necessary to go to a Bayesian model to find a strategy that is better than just accepting the first value. Indeed, we now present a strategy whose expected return is larger than $(A+B)/2$.

Let $P(x)$ denote any strictly increasing distribution function (for instance, the standard normal). Employ the following strategy: If the value of the check is x, then accept it with probability $P(x)$; that is, choose a random number U and accept x if $U < P(x)$. The expected return is

$$\begin{aligned} &E[\text{Return} \mid A]/2 + E[\text{Return} \mid B]/2 \\ &\quad = \{AP(A) + B[1 - P(A)]\}/2 + \{BP(B) + A[1 - P(B)]\}/2 \\ &\quad = \{A + B + [B - A][P(B) - P(A)]\}/2 \\ &\quad > (A+B)/2 \end{aligned}$$

where the inequality follows since $[B - A]$ and $[P(B) - P(A)]$ both have the same sign.

Since the value x will be accepted if $U < P(x)$, which is equivalent to $x > P^{-1}(U)$, it follows that the type of strategies given above are equivalent to ones that start by choosing a value Y from any specified continuous distribution function on the whole real line (assuming that A and B could be negative; if you know they are positive, then Y should be chosen from a positive continuous distribution) and then accepting the first value if and only if it is greater than Y. (If a deterministic value of Y is used, then the expected return will exceed $(A+B)/2$ if the value chosen falls between A and B, and it will equal $(A+B)/2$ otherwise. By letting Y be the value of a continuous random variable there will be a positive probability that it falls between A and B and will thus give rise to an expected return strictly greater than $(A+B)/2$.)

Another way of seeing that always accepting the first value cannot be a good strategy is by recalling that the only admissible strategies are ones that are either uniquely Bayes (for some prior on A, B) or the limits of such.

Sheldon M. Ross
IEOR Department
University of California
Berkeley, CA 94720

The deterministic rule appeared in our paper, suggested by a referee. Ross's analysis makes a nice addition to the problem.

Ronald CHRISTENSEN
Department of Mathematics and Statistics
University of New Mexico
Albuquerque, NM 87131

Jessica UTTS
Division of Statistics
University of California
Davis, CA 95616

SPROTT, D. A., AND FAREWELL, V. T. (1993), "THE DIFFERENCE BETWEEN TWO NORMAL MEANS," *THE AMERICAN STATISTICIAN*, 47, 126–128: COMMENT BY BERGER AND BOOS AND REPLY

Sprott and Farewell (1993) discuss the method of Barnard (1984) for making inferences in the Behrens–Fisher problem using $t(\delta;\rho)$ (their eq. (3)). This latter statistic has a t distribution with $m+n-2$ degrees of freedom under H_0: $\delta=\theta_1-\theta_2=0$ for a specified value of $\rho^2=\sigma_2^2/\sigma_1^2$. Sprott and Farewell suggest the reasonable and intuitive approach of looking at p values from $t(\delta;\rho)$ over a range of ρ, that is, over a 90% confidence interval for ρ.

We would like to point out, however, that $t(\delta;\rho)$ and the resulting inferences can be overly sensitive to ρ as compared to competing procedures. In their Example 1, the two-sided p value goes from .009 to .002 as ρ ranges over the 90% confidence interval (.65, 1.91). In their Example 2 the p value goes from .006 to .145 as ρ ranges over (.53, 3.30). Thus inferences in Example 2 are quite sensitive to the value of ρ and less so in Example 1.

An alternative approach is to use a statistic such as Welch's t, $t_w=(\bar{x}-\bar{y})/\sqrt{s_1^2/m+s_2^2/n}$, and to compute the p value $p(\rho)$ based on independent $N(0,\sigma^2)$ and $N(0,\rho^2\sigma^2)$ samples. This latter computation is nontrivial but easily carried out by Monte Carlo simulation. Details are in Berger and Boos (1994).

For Example 1 $p(\rho)$ goes from .0026 to .0012 as ρ ranges over (.65, 1.91). In Example 2 $p(\rho)$ goes from .0081 to .0344 over the range (.53, 3.30). (Estimates of $p(\rho)$ were based on 100,000 Monte Carlo replications and have standard errors in the fourth decimal place.) Thus $p(\rho)$ is more sensitive to ρ in Example 2 than in Example 1, but in either case inferences are less sensitive to ρ than for the analogous procedure based on $t(\delta;\rho)$.

The implied principle here is that inferences based on statistics such as $t(\delta,\rho)$ that depend directly on ρ are more sensitive to ρ than are analogous inferences based on statistics such as t_w where dependence on ρ only enters through the distribution of the data.

Roger L. BERGER
Dennis D. BOOS
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

REFERENCES

Barnard, G. (1984), "Comparing the Means of Two Independent Samples," *Applied Statistics*, 33, 266–271.

Berger, R. L., and Boos, D. D. (1994), "P Values Maximized Over a Confidence Set for the Nuisance Parameter," *Journal of the American Statistical Association*, 89, 1012–1016.

We welcome the additional consideration given by Berger and Boos to the topic discussed in our paper. The examination of alternative procedures serves to highlight important characteristics of the problem.

The main thrust of their letter, however, seems to rest on the subjective opinion that $t(\delta;\rho)$, and the resulting inferences, can be overly sensitive to ρ as compared to competing procedures. A competing procedure is then described, but no further justification appears to be given. Doubtless many other such competing procedures can be produced.

On this evidence, one could equally conclude the opposite, namely, that competing procedures and their resulting inferences can be unduly insensitive to ρ as compared with $t(\delta,\rho)$. No logical or statistical principle is put forward to support either view.

In this regard, the statement in the fifth paragraph of the comment is not an implied principle. It is an assertion that is either true or false. The logical principle that can be inferred seems to be that tests that have the property of similarity are optimal. But, as is well known, the requirement of similarity does not of itself lead to reasonable procedures.

David A. SPROTT
Vernon T. FAREWELL
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario
Canada N2L 3G1

THE MEAN IS WITHIN ONE STANDARD DEVIATION FROM ANY MEDIAN

O'Cinneide (1990) recently gave another proof of an old result—Hotelling and Solomons (1932)—that $|m-\mu|\leq\sigma$ for any median m of a distribution with mean μ and variance σ^2. This started a spate of letters (see Murty 1991; David 1991; Mallows 1991; Dharmadhikari 1991; and O'Cinneide 1991). We wish to give a proof, seemingly more elementary and intuitive than any of those offered. However, we think O'Cinneide's proof is very pretty and do not see why he should withdraw it!

Suppose without loss of generality that $\Pr(X\leq m)=\Pr(X>m)=1/2$ and that $\mu>m$. We want to change the distribution of X so that μ is kept constant but σ^2 minimized. Clearly σ^2 will be reduced if all the probability for x less than or equal to m is moved up to m and so closer to μ. Then the expectation of X is

$$m/2+\int_{x>m} x\,dF=\mu.$$

Denoting $2\int_{x>m} x\,dF$ by μ_+, the mean of the distribution above m, $\mu_+-\mu=\mu-m$. Writing $x-\mu=x-\mu_++(\mu_+-\mu)$, the contribution to σ^2 of the distribution above m is

$$\int_{x>m}(x-\mu)^2dF=\int_{x>m}(x-\mu_+)^2dF+(\mu_+-\mu)^2/2.$$

Thus σ^2 is minimized if the distribution above m is concentrated on μ_+. The variance is now $(m-\mu)^2/2+(\mu_+-\mu)^2/2=(m-\mu)^2$. Hence we have the desired inequality $\sigma^2\geq(m-\mu)^2$.

Exactly the same argument establishes the generalization

$$\sigma^2\geq(\zeta_p-\mu)^2\,\max\{p/q,q/p\},\tag{1}$$

where ζ_p is the p quantile of X and $q=1-p$. Actually it shows that if, for example, $\zeta_p\leq\mu$, then $\sigma^2\geq(\zeta_p-\mu)^2p/q$, which allows a smaller σ^2 than (1)—but you have to know more to benefit from it.

If one goes back to Hotelling and Solomons's article (1932), one finds that the proof is for sample estimators. The same volume contains another proof by Garver (1932), also for data. They both imagine that the median of n (odd) distinct observations $x_1,\ldots,x_n$, is zero and that $\bar{x}>0$ and define the standard deviation as $(\text{sd})^2:=\Sigma(x_i-\bar{x})^2/n$. (We note parenthetically that David's sample version of (1) would look nicer with this sd. It could be proved by the sample version of my proof of (1).) Then the skewness measure is called $s=\bar{x}/\text{sd}$ and their aim is to show that $s^2\leq1$. Hotelling and Solomons's proof is very like mine. Garver reduces $s^2\leq1$ to $n\Sigma x_i^2/(\Sigma x_i)^2\geq2$. Let $n=2m+1$ and call the m negative x_i's "y_i" and the m positive x_i's "z_i". Then Cauchy gives $n\Sigma y_i^2/(\Sigma y_i)^2\geq1$ and $n\Sigma z_i^2/(\Sigma z_i)^2\geq1$, so combining $m\{\Sigma x_i^2+\Sigma z_i^2\}/\{(\Sigma x_i)^2+(\Sigma z_i)^2\}\geq1$. Thus $n\{\Sigma x_i^2+\Sigma z_i^2\}/\{(\Sigma x_i)^2+(\Sigma z_i)^2\}\geq2$. But $n\Sigma x_i^2/(\Sigma x_i)^2$ equals $n\{\Sigma x_i^2+\Sigma z_i^2\}/\{\Sigma y_i+\Sigma z_i\}^2$, so the result follows since the product $\Sigma y_i\,\Sigma z_i$ is negative. Garver then allows for equal observations and for n even. Hotelling and Solomon are less general and less convincing. One final historical note: Anyone interested in handwriting will find the signatures of the authors of papers in these old volumes. It seems, however, that the same hand wrote out *all* formulas in *all* papers!

This work was supported by National Science Foundation Grant DMS 9212415.

G. S. WATSON
Department of Mathematics
Princeton University
Princeton, NJ 08544-1000

REFERENCES

David, H. A. (1992), "Mean Minus Median: A Comment of O'Cinneide," *The American Statistician*, 45, 257.
Dharmadhikari, S. (1991), "Bounds on Quantiles: A Comment on O'Cinneide," *The American Statistician*, 45, 257–258.
Garver, R. (1932), "Concerning the Limits of a Measure of Skewness," *The Annals of Mathematical Statistics*, 3, 358–360.
Hotelling, H. O., and Solomons, L. M. (1932), "The Limits of a Measure of Skewness," *The Annals of Mathematical Statistics*, 3, 141–142.
O'Cinneide, C. A. (1990), "The Mean is Within One Standard Deviation of Any Median," *The American Statistician*, 44, 292–293.
Mallows, C. (1991), "Another Comment on O'Cinneide," *The American Statistician*, 45, 257.
Murty, V. N. (1991), "A Comment on O'Cinneide," *The American Statistician*, 45, 257.

Following the impact of texts by Tukey and Ehrenberg, most of us now teach our students not to clutter up tables with distracting detail; but the subeditors and compositors, who prepare our journals, are not always aware of this objective.

May I therefore first congratulate *The American Statistician* on being one of the few journals (in any discipline) that does not insert leading zeros, when these have been suppressed by an author. Historically, with the manuscript decimal-period often at risk of getting overlooked, the initial zero was a sensible precaution. In modern print, this is hardly a danger and is quite superfluous—especially in tables. I do not find it helpful, for instance, to have a vector (.1, .2, .3) replaced by (0.1, 0.2, 0.3).

On logical grounds, I also object to certain punctuation conventions that are practised (and, here, this publication fares less well). The one most prevalent is illustrated by the majority of references and the heading of almost every published letter to the editor, such as '... "On Generalised Score Tests,"...', rather than '..."On Generalised Score Tests", ...'.

No way should a quoted title include a terminating comma. The absurdity of this practice is emphasised in certain contexts. For instance, a sentence in the *typeset* rejoinder from Huberty (1993, p. 314) appears as: 'The same is true for "ordinal data," "interval data," and "ratio data."' It is interesting to note the difficulty that the compositor had over the less familiar semicolon in the same letter—we get the sensible '"ordinal data";' and '"... interval scale";' mixed with the throw-back of '"ordinal variable;"'.

When the quotation ends with a question mark, sub-editors realise that to try '"Quo vadis?," a quotation from the Vulgate.' is truly ridiculous, and so sacrifice punctuation and go for '"Quo vadis?" a quotation ...', when of course the commonsense treatment gives '"Quo vadis?", a ...'.

The misuse of the abbreviating period is another node of contention. What is the sense of printing 'Dr.' or 'Jr.'? Nothing has been truncated after the 'r'. Either follow the lead indicated by '*J. Opl Res. Soc.*', which distinguishes between the cases of truncation and internal contraction (without including an overstudied and self-conscious omission's apostrophe) or just observe the (good) modern business practice of suppressing much of the conventional punctuation in correspondence (such as dropping trailing commas, and not indenting paragraphs).

Similarly, 'Vol. 1' and 'Vols 1&2' are correct, but 'No. 1' gives a problem. Of course, the business convention of suppressing the period removes the difficulty, with 'Vol 1' and 'No 1' then being written. But 'No. 1' is emphatically wrong, as 'No' is the contraction of the Latin ablative 'Numero'; although, to emphasise this, the continental European contraction, 'n°', provides a rather more elegant option.

We are right to criticise so-called educated folk who proudly boast their lack of numeracy. Let us be aware that ignorance of the norms set by the literate is no less moronic. But, may we please temper poorly conceived compositing conventions with logical precision?

Oliver D. ANDERSON
Department of Statistical and Actuarial Sciences
University of Western Ontario
London, Ontario
Canada N6A 5B7

The punctuation style advocated by Anderson and used in his letter is called "British style" by American publishers. The ASA uses American-style punctuation as well as American spelling in its journals.

The ASA recommends to its authors *The Chicago Manual of Style*, a style guide widely accepted among publishers, editors, and universities: "The British style is strongly advocated by some American language experts. In defense of nearly a century and a half of the American style, however, it may be said that it seems to have been working fairly well and has not resulted in serious miscommunication ... [T]he University of Chicago Press continues to recommend the American style for periods and commas" (p. 161).

The inconsistent placement of the semicolon cited by Anderson was likely a typographical error. I know of no style guide currently in use by the ASA or anyone else that recommends including semicolons in quotation marks.

As with commas and quotation marks, current use of the abbreviating period in ASA journals seems to be familiar and easily understood by our readers and authors.

Rich DODENHOFF
ASA Journals
Alexandria, VA 22314-3402

REFERENCE

The Chicago Manual of Style (14th ed.) (1993), Chicago: University of Chicago Press.

CORRECTION

C. J. Wild and G. A. F. Seber (1993), "Comparing Two Proportions From the Same Survey," *The American Statistician*, 47, 178–181.

On page 179, in the equation at the bottom of the left column, $\hat{p}_{12} - \hat{p}_{22}$ should be $\hat{p}_{12} - \hat{p}_{21}$. In the first line of Equation (2), var$[\hat{p}_1 - \hat{p}]$ should be var$[\hat{p}_1 - \hat{p}_2]$.

On page 180, in the left column, the second displayed equation should be

$$p_{12} + p_{21} = p_1 + p_2 - 2p_{11}.$$

In Equation (3), $\min(p_1 + p_2,\ p_1 + q_2)$ should be $\min(p_1 + p_2, q_1 + q_2)$. In the second full paragraph of the right column, $\hat{p}_1 + .005$ should be $\hat{p}_1 \pm .005$.

On page 181, Section 3, the numerator in the second display equation should be $|\hat{p}_1 + \hat{p}_2|$, and the denominator should have min instead of max.

The authors thank John Harder for finding these errors.