STATISTICAL LITERACY

L. Mike Perry, Appalachian State University

Introduction

In our contemporary society, widespread awareness of information technology has become increasingly important for public understanding of day-to-day affairs as well as national policy-making. Many observers believe that if the gap between current technological capacity and the ordinary person's understanding of that technology is allowed to grow too wide, the social, cultural, political, and scientific ramifications are likely to be quite serious (Molnar, 1981).

The importance of public understanding of computers has prompted concerned educators and civic leaders to urge massive efforts to make "Computer Literacy" the fourth "R", with major curriculum changes proposed for our schools and colleges. A plethora of articles on "Computer Literacy" have been forthcoming and hardly a day goes by without some reference in the local newspaper to computing in the schools. Public support for computer literacy seems high. The same does not appear to be the case for "Statistical Literacy" and yet statistical understanding in an information society may be even more important.

Dictionary definitions of language literacy include (1) the ability to communicate, i.e. reading and writing and (2) the state of being informed, "cultured", and well versed. "Literate" in this context sometimes means a minimal level of literacy and in other cases means a very high level of literacy.

Early definitions of scientific and technological literacy suggest that the term literacy is usually intended to mean only a matter of being informed rather than requiring higher level thinking and communication skills. Scientific and technological literacy (Pella, O'Hearn, and Gale, 1966) were defined as scientific and technical knowledge pertaining to social implications. Thus the science-society interface has played a prominent role in most definitions. In many instances "Science/Technology and Society" has been equated with the term "Scientific and Technological Literacy".

A definition of "Statistical Literacy" should include

A. Statistical thinking

B. Communicating statistically

as well as

C. Knowledge

D. Social implications

This paper will concentrate on thinking and communicating. No attempt, however, is made here to define a "knowledge syllabus" of conceptual and philosophical understandings, methodological or problem solving skills, or required studies of societal impacts for a course. Previous discussion on theses issues has been extensive, including articles by Ehrenberg (1976)', Kempthorne (1980), Kerridge (1976), and Tukey (1980). Moreover, the. author is not necessarily proposing a new course in "Statistical Literacy" but is strongly suggesting desirable objectives for an undergraduate introductory statistics course.

Experiences with students referred to later in this paper were encountered in a required undergraduate course for social science majors. This course serves as both an introduction to and a termination of statistics for most of these students! Pressure exists from the social science departments to teach their students "all the statistical research tools they will ever need" in one semester, an assignment' which is, of course, impossible and leads to an overwhelming syllabus of topics and techniques. Hereafter, 'introductory course" refers to a statistics course of this nature.

First Encounters with Ambiguity

On many campuses, a majority of the students who populate the introductory statistics class have what we as professors perceive to be (1) considerable fear and anxiety about the subject, (2) "academic weaknesses", and (3) lack of motivation for learning (other than perhaps the threat of a bad grade). These are difficult for the best of us to overcome. "Academic weaknesses" are not remedied by prerequisites. Arons, a physicist,' has suggested (1979) that assumptions made by professors about student reasoning skills are too often unreasonable. Alternate approaches and revised objectives are too often and easily criticized as weakening the course at the expense of the syllabus. A framework for considering "statistical thinking and communicating" is necessary.

Most students enter the introductory statistics course with the notion that it is another mathematics course. In some cases, unfortunately, they are correct. The student's concept of mathematics is typically limited to college or high school algebra. Unambiguous questions are posed which always have answers (usually unique or at worst a finite number of solutions which can be logically validated). If the problem is: $2X + 1 = 5$ then the answer is: $X = 2$! If the algorithm is followed and the arithmetic is performed correctly then the probability is one that the answer is correct. Numbers represent fact and precision!

"Statistical thinking" is a first encounter with ambiguity in a context with numbers. A data problem presented in the form "Explore this data and see what can be found" is not perceived to be a "problem". There are no "questions"! Given a "question" it is a startling revelation that (1) this may not be the proper question, (2) it may not have an answer, (3) there may be more than one answer, each valid, but several seemingly contradictory, (4) the answer cannot be validated, (5) the investigator never knows if all the relevant questions have been asked, etc. The concept of variability of estimation is crucial but elusive. Interpreting the results of an inferential procedure is another shocker. What good is an answer if (1) it probably is in error (even if the possible error is estimated) and (2) there is a chance (however small) the answer is incorrect. What about a '!good" fit? It depends on which criteria you use! And how do you "criticize" an experimental design if there is no recipe for criticism.

"Statistical thinking" requires dealing with ambiguities and open ended situations. This is a significant leap for the student from the methodological processing of algebra. It is one of the characteristics of the subject which distinguishes it from being "just another math course." The nurturing (teaching?) of "statistical thinking" must be an objective of the introductory course.

Next we will examine the notion of "intellectual maturity", distinguish it from "intelligence", and relate it to "statistical thinking" and ultimately to "communicating statistically".

Intellectual Maturity

William Perry (1970) and his associates at Harvard University developed a theory which outlines the intellectual and ethical development of college students. This nine-position scheme traces the evolution in students' thinking about the nature of knowledge and values. The scheme describes the steps by which students move from a simplistic, categorical view of the world (described in such unqualified terms as good-bad, right-wrong, black-white) to a realization of the contingent nature of knowledge, relative values, and the formation of their commitments. It addresses the interface between their intellect and the way they perceive knowledge. Heffernan (1975) gives a straightforward discussion of the Perry model.

Levels of Intellectual Development (William Perry)

- Dualism (1,2)
  - Knowledge is absolute, teacher is authority
  - All questions have right/wrong answers
  - The good student masters these answers
  - The good teacher teaches only the right answers
  - Tasks with options are confusing
  - Different points of view are frustrating
- Multiplicity (3,4)
  - Acknowledge multiple perspectives for a given problem, but accept all points of view equally valid and not subject to evaluation
  - Begin to see difference in considered belief and an unconsidered judgment
- Relativism (5,6)
  - Capacity for analytic thinking and judgment All knowledge and values seen as relativistic and contextual, But commitment to choices from "multiple truths" are difficult
- Commitment in Relativism (7,8,9)
  - Attain intellectual and ethical maturity
  - Full affirmation of choices is made
  - Knowledge is a personal responsibility

The William Perry development scheme has undergone considerable scrutiny. King (1978) gives an extensive review of basic research on Perry's theory. Blake (1976) questioned whether students in a science-oriented curriculum would show an upward trend over the college years. His results showed a significant movement toward maturity. The Perry scheme has proved to be a highly useful tool for practitioners both' in understanding students and designing programs to promote their development. Copes (1980) points out the pertinence of the model as a useful framework for college science and mathematics teachers.

The nurturing of "Statistical thinking" in the introductory statistics course involves pushing the student up the Perry hierarchy. The dualistic level that may suffice to solve a simple alegbra problem with a unique numerical solution is not sufficient to tackle the open-ended analysis of data. The student who can successfully criticize an experimental design or understand the interpretation of statistical inference must have matured to relativism. The ambiguities presented by statistical problems can only be confronted by the student who has a high level of maturity.

Communicating statistically

Effective communication of statistical ideas involves presenting numerical and graphical descriptors and summaries along with concise verbal summaries and interpretations. It requires the integration of thinking and reasoning skills with these media to convey concepts and ideas in a coherent fashion. The consequences of thinking are not apparent without effective communication. Conversely, developing effective communications skills requires development of reasoning skills. Hays (1982) and others have pursued this theory. Thus effective "Statistical communication" is a desirable objective for the introductory course.

The students, confronted for the first time with numerical "ambiguity", must now integrate the numerical, the graphical, and the verbal forms coherently! They may have previously succeeded in responding to a specific request to draw a bar-graph, compute a median, or to discuss in a paragraph the definition of a statistical term. A significant increase in maturity, however, is required to respond to an open-ended question which requires the concurrent use of thinking skills and several communication forms.

Teaching Considerations

Encouraging "Statistical thinking" and effective "Statistical communication" in an introductory statistics class is not an easy task. It requires considerable patience on the part of the professor! It requires changes in the nature of assignments and test questions. It demands a change in the approach to grading--in the early stages of the course objective criticism without devestating penalty must be provided. It requires more time for open-ended classroom discussion than is usually provided. It requires more time than is normally demanded for critique of students' papers. Appropriate texts and materials may be difficult to find. Books such as those by Freedman et al (1978) or Moore (1979) provide nice presentations of concepts and controversies but may need supplementing to cover other objectives. The reaction of "dualistic" students who expect the professor to be the "authority" may even be disruptive when they are confronted with ambiguous assignments. Critiques of these assignments which do not (in their required terms) explain why their responses are not one hundred percent sufficient frequently produce hostility. This is especially the case when the professor cannot (interpreted as will not) provide the correct answer.

Strategies

I have used two types of assignments in the introductory course to encourage statistical thinking and communication--(1) projects and (2) discussion questions on tests and examinations. A project, for example, might involve an open-ended analysis of a reasonably large data set, be assigned over a three week time period, and result in a written report. Projects are discussed in another paper by Gary Kader in these proceedings.

Good questions for a one-hour exam which test for statistical thinking and communication skills are somewhat difficult to devise; response time is usually limited to no more than twenty minutes. The examples presented here are not trivial but experience suggests that they are appropriate for this time limit.

The first part of my introductory course covers descriptive statistics and exploratory data analysis. Thus on Test 1 two questions of the following nature might be desirable:

Question #1 - Give a small data set. Ask the student to compute the quarter summary, fences, outliers, etc.

Question #2 - Describe a "situation" in terms of quarter summaries. Ask the student to "criticize".

Question #1 does not test for statistical thinking and communication. It is a "math" problem, requires correct application of algorithms and accurate arithmetic, is not ambiguous, the answers are unique, and the students feel comfortable with it. It tests for facility with computation and knowledge of the definitions. They won't, of course, all get it correct!

Question #2 requires statistical thinking and the students must communicate their results. It may present ambiguities and several possible "correct" answers. A question of this type with examples of student responses follows.

Example test problems

Question #2 - Based on the following summary of scores on a test (scale 0-100), the claim was made that females did considerably better than males by noting that the male median score is significantly higher (10 points in fact). Criticize this claim.

Summary of scores

	MIN	QI	MEDIAN	Q3	MAX
Female	48	60	82	86	90
Male	62	68	72	92	97

Note--there are no outliers in either group each group contains 100 persons.

Some Responses to Question #2:

Student A:

	Females	Males
Lf=	1.5	32
Uf=	144.5	128
IF= '	-18	-4
UF=	164	164
Min=	48	62
Max=	90	97
MidQ	73	80
MidR=	474	79.5
Q-Sprd=	26	24
Med=	82	72
Mean=	73.2	78.2
Range=	42	35

"The range for females is higher" (no graph is shown)

Student B: "The median is found by putting the numbers in ranking order. The median is not affected by extreme values in either direction. This is why it is important to give the other data and by observing it we can tell that the males did better on the exam." (an accurate graph is shown'

Student C: "The median is just a number representing the score that is in the middle of all the scores. There could be a lot of low scores on the test for females and a lot of high scores: The number picked could be just an odd score that happened to be in the middle. The same would be true of the males. Thus, a better means of computing the scores would be the average or mean."

Student D: (She drew the two boxplots accurately and then commented) "One cannot judge distributions accurately by simply examining the median. I find that the claim is easily proven false once the data is visualized in a boxplot drawn on the scale of 0-100. Although the median of female scores is higher it is shown here that overall the males scored significantly higher than the females by observing in the graphs that the top quarter for the males is higher than the top quarter for the females. Also the bottom quarter for the females is lower than the bottom quarter for the males.

Student E: "The males had 25% of their top grades between 92 and 97 while the females had the top 25% between 86 and 90; the lower 25% of grades for males was between 62 and 68 while they were between 48 and 60 for the females. The midQ and mid-R are higher for males than females also." (an accurate graph of boxplots is shown)

Student F: "The distribution of scores has a greater range with the males. The males, thusly, had the greater range. The results could be biased for one group or the other. There's much more variance among the males so it really wouldn't be fair to say the girls did 10 points better. This summary is not valid because it doesn't measure true scores, just the median, and how the scores are dispersed. I would say as a whole that males did better. The test is reliable because it can be done like this over and over again and you'll get the same results, but these results aren't really what we're looking for." (There is a drawing of the boxplots but it is not uniformly scaled, is out of proportion, and misleading)

Student G: "The median is the middle number with respect to the order of this group of numbers. In this summary we took the middle number to base our opinion not the actual average of the scores. The median is not affected by extreme values in either direction. The average would be affected by extreme values in either direction so I would think that it would be more applicable here instead of the median. I would say that in figuring house prices the median would be better to use because you do not get this high, low distortion. But in figuring these numbers I think the mean would be more accurate. If the average would have been taken the males would show a higher mean. This is another reason why it is important to observe other data and be sure that your first explanation is not the only explanation." (no graph shown)

Student H: Acceptable graph is shown but no comments.

Comments on Responses:

Student A: Dualistic--No thinking or communicating. Has not "looked" at the data but sees the questions as a "math" problem and proceeds to compute every numerical quantity that may be related but does not relate these to the situation. Makes a verbal comment because knows it is required. Does not use graph as thinking or communicating device.

Student B: Dualistic--A verbalizer. Uses no numbers or graphs. Has not "looked" at the data. Answers to questions come from his notes; thus proceeds to recite a definition and comment on what should be done to respond properly but does not "observe" anything.

Student C: Dualistic--Similar to B. Remembers that under some circumstances the mean might be different from the median (from the notes) but cannot "look" at the data and apply this notion in context.

Student D: Has used the graphical display effectively as a thinking and communicating device. Has "looked" at the data and conveyed verbally what she has seen. Does not, however, use specific numbers to describe upper and lower quarters in her comments.

Student E: Similar to D but uses specific numbers in comments.

Student F: Dualistic--Responds from notes but chooses wrong section to begin. Comments on bias, reliability, and validity are "echos".

Student G: Dualistic--Recites from notes. Suggests that the mean should be higher for males, but like C he is not convincing.

Student H: Dualistic--Remembers that the graph has something to do with quarter summary (it comes next!). Can draw the boxplots accurately but cannot interpret in context.

Summary of example: The boxplot is a simple but powerful device for looking at data; when used with precaution it is a useful thinking and communicating schema. The results of this example for 76 students seem to bear this out. A large percentage of acceptable responses utilized an acceptable boxplot and vice versa as suggested by the following summary data. (A=acceptable, N=not acceptable)

Boxplot			
	A	N	
A	37	7	44
	.48	.09	.58
	.84	.16	
	.77	.25	
Response			
N	11	21	32
	.14	.27	.42
	.34	.66	
	.23	.75	
48	28	76	
.63	.37		

Another Example

Here's a short problem of this nature to test another area for "Statistical thinking and communicating". For inference the typical problem is:

Question #1 - Give some sample data. Ask to construct a 95% confidence interval.

A dualist can handle this but the following requires more.

Question #1.5 - Interpret in context the result of #1.

Next is the problem to test for the non-dualistic thinker. It does not present specific questions, confronts the respondent immediately with ambiguity, and requires only slight use of arithmetic. The graphical aspects are simple--pictures of intervals. It gets to the essence of the basic concepts of sampling variability and confidence interval without using these terms.

Question #2 - Two polling firms conducted a poll on the same day, each using a random sample size of 400 persons. Both firms used sound sampling procedures and well thought out survey methods which could not be criticized.

However, the results of the surveys seemed quite different. One showed that 44% of the American public favor the President's economic policy but the other indicated considerably less --only 36% support the president's policy.

As a statistician you are asked to explain: (a) How we can get two valid answers to the same question, and (b) How the apparently different answers are not so different as they may at first appear and there is no contradiction in the results of the two polls.

Summary

The encouragement of "Statistical thinking and communicating" must become a key objective of the introductory statistics course if we desire to have a truly "Statistically literate" public. The William Perry maturity scheme offers a useful framework for implementing strategies for accomplishing these goals. The concept of moving the student up the hierarchy away from dualism seems especially relevant as a criterion in defining course objectives.

Note--This paper is in part a consequence of the project "Statistical Experiences for Undergraduates" supported by the National Science Foundation's Local Course Improvement Fund Grant #SER-8161061 in the Department of Mathematical Sciences, Appalachian State University, 1981-83.

References

Arons, A.B. "Some Thoughs on Reasoning Capacities Implicitly Expected of College Students." In J. Lockheed and John Clement (Eds.), Cognitive Process Instruction-Research on Teaching Thinking Skills. The Franklin Institute, 1979.

Blake, L. "A Measure of Developmental Change: A Cross-Sectional Study." Paper presented at 84th annual meeting of the American Psychological Association, Washington, D.C., September 1976.

Copes, L. "College Teaching, Mathematics, and the Perry Development Scheme." Unpublished paper, Institute for Studies in Educational Mathematics, St. Paul, Minn., 1980.

Ehrenberg, A.S. "Symbols or Concepts?" The Statistician, 25, 3, 191-193, 1976.

Freedman, D; Pisani, R.; and Purves, R. Statistics, New York: W.W. Norton, 1978.

Hays, J. "Development of Analytical Writing Ability: Preliminary Report." Institute for Studies in Educational Mathematics, St. Paul, Minn., 1982.

Heffernan, J. "An Analytical Framework for Planning and Research in Higher Education." Liberal Education, 1975,61,493-503.

Kempthorne, 0. "The Teaching of Statistics: Content Versus Form." The American Statistician, 34, 1, 17-21, 1980.

Kerridge, D. F. "The Menace of Mathematics. "The Statistician, 25,3, 179-189, 1976.

King, P.M. "William Perry's Theory of Intellectual and Ethical Development." In New Directions for Student Services: Applying New Developmental Findings, no. 4. San Francisco, Jossey-Bass, 1978.

Molnar, A. R. "The Coming of Computer Literacy: Are We Prepared for It?" Educational Technology, 26-28, January 1981.

Moore, D.S. Statistics-Concepts and Controversies. San Francisco: I. H. Freeman, 1979.

Pella, M.O.; O'Hearn, G.T.; Gale, C.W. "Referents to Scientific Literacy." Journal of Research in Science Teaching, 4, 199-208,1966.

Perry, W.G., Jr. Forms of Intellectual and Ethical Development in the College Years: A Scheme. New York: Holt, Rinehart and Winston, 1970.

Tukey, J. "We Need Both Exploratory and Confirmatory." The American Statistician, 34, 1, 23-25, 1980.