QUESTIONS TO ASSESS THE UNDERSTANDING OF STATISTICAL CONCEPTS

John D. McKenzie, Jr.
Babson College, United States
Robert Goldman
Simmons College, United States
mckenzie@babson.edu

The GAISE College Report recommends that introductory applied statistics courses should place greater stress on statistical concepts and less stress on definitions, computations, and procedures. The report also urges instructors to align assessments with learning goals. In this paper the authors explain how instructors can implement these two recommendations. They first review the extent to which questions directly related to concepts are found in popular texts and on websites created by the publishers of these texts. Following this review they provide examples of such questions in a variety of formats (multiple choice, fill-in-the blank, open-ended, etc.). The examples will be classified by the approximate level of educational objective contained in Bloom's Taxonomy. Finally, the paper will discuss the advantages and disadvantages associated with having students answer such questions electronically.

INTRODUCTION

In this paper the authors will discuss how instructors, who teach the introductory applied course in statistics, can introduce questions to assess a student's understanding of important statistical concepts. The Guidelines for Assessment and Instruction in Statistical Education (GAISE) College Report (2005) suggests reasons for the introduction of such questions. First, the report recommends that introductory applied statistics courses should place greater stress on core statistical concepts and less stress on definitions, computations, and procedures. Second, the report also urges instructors to align assessments with learning goals. Of course, the understanding of concepts should only be assessed if they are taught.

Such assessment may be implemented in numerous ways, but the most common approach is to use questions on examinations and quizzes. Cobb (1987) suggests that the most important criterion in selecting a textbook is the quality of the questions. Such questions may take numerous formats such as multiple choice, fill-in-the blank, and open-ended. In this paper we will discuss the availability of conceptual questions from publishers and other sources. We will also present examples of such questions (and non-conceptual questions) based upon Bloom's taxonomy. In conclusion, there will be a discussion of electronic assessment.

TEXTBOOKS

In this section, we summarize our search for questions dealing directly with core statistical concepts in six introductory applied textbooks written by best-selling authors in the United States. We also comment on the questions available in these texts' supplements (e.g., test bank supplements in print and on the web). Three of these textbooks are marketed for general education courses. These are De Veaux, Velleman, and Bock (DVB) (2006), Johnson and Kuby (JK) (2004), and Triola (T) (2004). Three are marketed for business and economics courses: Anderson, Sweeney, and Williams (ASW) (2002), Berenson, Levine, Krehbiel (BLK) (2006), and Moore *et al.* (MMDS) (2003). McKenzie (2004) points out that while there are no agreed upon definitions of "core concepts" in statistics most statisticians would agree that "variability," "association," and "sampling distributions" are important concepts. In order to obtain a sense of the relative quantity of conceptual questions in the six texts, we reviewed the questions in the chapters in which sampling distributions and association were introduced. We defined a conceptual question as one in which the direct focus is on the student's understanding of a statistical concept or how the concept is operationalized in practice; it may follow on the heels of a computation or another non-conceptual question. Non-conceptual questions are primarily computational, definitional, or procedural.

Each question on association (A) and sampling distributions (SD) in each text was classified by whether it was conceptual, non-conceptual, or one that contained a mixture of both types. The results are set out in Table 1 below.

Table 1: The Number of Each Type of Question by Textbook

	General Education						Business and Economics					
	DVB		JK		T		ASW		BLK		MMDS	
	A	SD	A	SD	A	SD	A	SD	A	SD	A	SD
Conceptual	14	0	5	0	6	2	0	0	2	1	15	8
Mixture	30	5	5	7	12	7	22	9	16	13	26	9
Non-Conceptual	23	39	17	21	15	29	1	28	0	20	0	17

The results are unambiguous but not perhaps surprising. The great majority of the questions are not conceptual in nature. Over the six texts only 53 or 12.5% of the questions were fully conceptual. "Mixture" questions invariably included a number of non-conceptual parts followed by a conceptual "conclusion". Here is a sampling of the conceptual questions:

- Data Set 4 in Appendix B includes a sample of 104 body temperatures of adults. If we were to construct a histogram to depict the shape of the distribution of that sample, would that histogram show the shape of a sampling distribution of sample means? Why or why not? (Triola (2004))
- Justify the statement "$\overline{x}$ becomes less variable as n increases." (Johnson and Kuby (2004))
- Explain why you could estimate the probability that 100 people selected at random had worked for their employers an average of 10 years or more, but you could not estimate the probability that an individual had done so. (De Veaux, Velleman, and Bock (2006))

In citing this relative dearth of conceptual questions we do not mean to disparage all non-conceptual questions. Indeed, many of these questions, particularly in DVB and MMDS, required considerable understanding of the underlying concepts in order to set up the calculations. Students need to practice the formulation of problems and both students and instructors benefit from a wide range of applications. Of course, the lack of balance in conceptual questions reflects the fact that, as instructors, we tend to focus on the non-conceptual aspects of statistics in our courses. Among the reasons for this imbalance are the facts that teaching concepts is difficult for instructors and mastering concepts is difficult for students. More conceptual questions in textbooks will surely come as instructors begin to focus more on concepts. There is certainly room for a greater number of conceptual questions in textbooks.

Virtually all the questions we reviewed in these six texts were open-ended; over all the 424 questions in the three general education texts the only non-open-ended question was a single multiple-choice question in Johnson and Kuby. Oddly, the various supplemental "test-banks" that accompany many of these texts contain a much greater variety of question types. For example, BLK's test item file has 932 multiple choice, 685 true/false, 692 fill-in, and 476 open-ended questions. Most of which were stand-alone questions; not part of a sequential set of questions.

It is not easy to compare the intrinsic difficulty of conceptual and non-conceptual problems but Bloom's taxonomy offers a starting point. The taxonomy, created by the psychologist B.S. Bloom and his colleagues, divides cognitive objectives into a hierarchy of categories. These six categories from lowest to highest are *knowledge* (the ability to recall information), *comprehension* (the ability to understand and/or translate the meaning of material), *application* (the ability to apply learned material to new contexts), *analysis* (the ability to break down material into component parts so that the whole may be better understood), *synthesis* (the ability to pull together diverse elements into a new structure), and *evaluation* (the ability to judge the value of ideas/material). Chance, delMas, and Rossman (2004) introduce the taxonomy in the context of designing different types of assessment tools.

In our judgment, few of the 424 questions we examined required only the lowest category (knowledge) and very few required the two highest categories of the taxonomy, synthesis, and

evaluation. A crude examination of the categories associated with a sample of the conceptual and the non-conceptual problems suggests that the non-conceptual problems fall, on average between the second and third categories, whereas the conceptual problems fall, on average, between the third and fourth categories of the taxonomy.

SOURCES OF CONCEPTUAL PROBLEMS

The instructor wishing to introduce a greater emphasis on statistical concepts in the first applied course has a number of sources that can provide models for conceptual questions. The appendix to the GAISE College Report itself contains suggestions for transforming simple, true-false questions into more thought-provoking multiple-choice questions. It also provides a further 37 innovative questions many of which are primarily conceptual. The ARTIST (Assessment Resource Tools for Improving Statistical Thinking) project has compiled an assessment tool consisting of 40 challenging multiple-choice questions. Virtually all of them focus on determining how well a student understands an important statistical concept. Researchers at the University of Oklahoma are developing a Statistics Concept Inventory designed to assess student's understanding of statistics. Murphy (2006) reported on the project, with examples of the conceptual problems that they have constructed, at the Joint Mathematics Meetings in January 2006. At the same conference Rossman and Chance (2006) provide examples of assessment questions "…to improve Student Learning in Introductory Statistics." Finally, it should be noted that the College Board's past AP statistics exams offer thoughtful questions many of which are primarily conceptual.

FOUR CONCEPTUAL QUESTIONS

We follow this brief survey of conceptual problems with several examples of conceptual questions that the authors have used in their own quizzes and exams. There is one example for each of the first four categories of Bloom's taxonomy. In our experience the two highest categories of the taxonomy—synthesis and evaluation—are rarely appropriate in an introductory course. The GAISE report emphasizes the need to stress (and assess) the understanding of statistical concepts. The questions below assess understanding of one of the most important and fundamental concepts in statistics, that of variation.

1. *Knowledge*

Circle the response you feel is the most appropriate.

One of the most important concepts in statistics is (a) the absence of variation (b) the lack of variation (c) the presence of variation (d) the omnipresence of variation.

2. *Comprehension*

Fill in the most appropriate word in the sentence below.

The inter-quartile range is a measure of the degree of _____ in a data set.

3. *Application*

This is a standard deviation contest. You must choose four numbers from the whole numbers 0 to 10, with repeats allowed.

(a) Choose four numbers that have the smallest possible standard deviation.

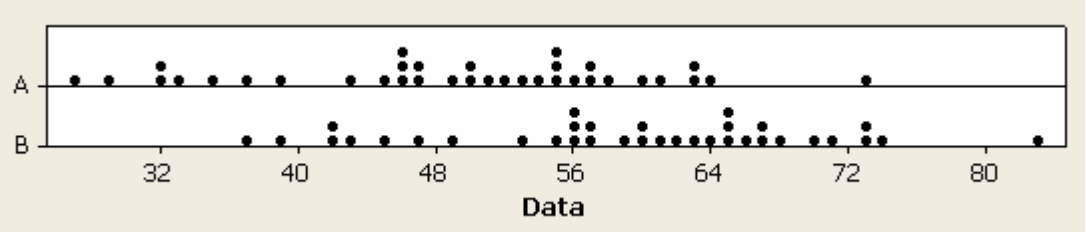
(b) Choose four numbers that have the largest possible standard deviation.

(c) Is more than one choice possible in either (a) or (b)? Explain.

4. *Analysis*

Dotplots of two integer data sets are shown below. The mean in data set A is 50 and the standard deviation 10. The mean in data set B is 60. What is the standard deviation in data set B?

(a) 12 (b) 20 (c) 10 (d) cannot be determined from the graph



ELECTRONIC ASSESSMENT

Many publishers in the United States allow the students in courses that have adopted their texts to access and answer questions electronically for quiz or examination credit, for homework, or for practice. Prentice-Hall's Grading Assist (PHGA) program is a good example. Most of these electronic questions are generated by two software vendors, Brownstone and Maplesoft. PHGA is generated by Brownstone's EDU engine. In addition, individual instructors have used such software to develop their own individual electronic questions.

Most of these programs randomize components of each question so that each time the program is run a slightly different version of the question appears with feedback on the student's response immediately available. There are benefits to using such software. One is that a student (or students) can work a question numerous times until the student or team of students is able to master the relevant material. Also, these programs provide a congenial environment to students, more used to playing video games than reading newspapers. This may enhance their willingness to spend time on their learning. Benefits to the instructor include the ability to use endless variations of well-designed questions easily, the availability of information such as the number of times a question has been attempted and the amount of time spent working a question by an individual student, as well as the grade on a question. Such programs reduce cheating on examinations and quizzes. Two disadvantages are the unwillingness of the software to provide partial credit and the need for the technology to implement the software. We conclude with some examples of electronic conceptual problems.

On a more personal note, each of the two authors reviewed the other's final exam from their introductory statistics courses. Both contained a few conceptual questions but the great majority were primarily non-conceptual. As a consequence of our research for this paper, we vowed to talk more about concepts in class and to introduce even more conceptual problems in our exams.

REFERENCES

Anderson, D. R., Sweeney, D. J. and Williams, T. A. (2002). *Statistics for Business and Economics* (8th Edition). Cincinnati, OH: South-Western.

Garfield, J., del Mas, R., Chance, B. and Ooms, A. (2005). *Assessment Resources Tools for Improving Statistical Thinking (ARTIST)*, https://app.gen.umn.edu/artist/.

Berenson, M. L., Levine, D. M. and Krehbiel, T. C. (2006). *Basic Business Statistics* (10th Edition). Upper Saddle River, NJ: Prentice Hall.

Chance, B., delMas, R. and Rossman, A. (2004). Designing and evaluating assessments for introductory statistics. Minicourse presented at the Joint Mathematics Meetings, Phoenix, AZ: Mathematical Association of America.

Cobb, G. W. (1987). Introductory textbooks: A framework for evaluation. A comparison of 16 books. *Journal of the American Statistical Association*, 82, 321-339.

De Veaux, R. D., Velleman, P. F. and Bock, D. E. (2006). *Intro Stats* (2nd Edition). Boston, MA: Addison-Wesley.

GAISE College Report. (2005). Guidelines for Assessment and Instruction in Statistical Education College Report. Alexandria, VA: American Statistical Association, http://www.amstat.org/education/gaise.

Johnson, R. and Kuby, P. (2004). *Elementary Statistics* (9th Edition). Belmont, CA: Duxbury.

McKenzie, J. D., Jr. (2004). Conveying the core concepts. *2004 Proceedings of the Statistical Education Section*, (pp. 2255-2257). Alexandria, VA: American Statistical Association.

Moore, D. S., McCabe, G. P., Duckworth, W. M. and Sclove, S. L. (2003). *The Practice of Business Statistics*. New York: W. H. Freeman and Company.

Murphy, T. J. (2006). The statistics inventory concept inventory: An instrument for assessing student understanding of statistics concepts. Paper presented at the Joint Mathematics Meetings. San Antonio, TX: Mathematical Association of America.

Rossman, A. J. and Chance, B. L. (2006). Examples of assessments to improve student learning in introductory statistics. Paper presented at the Joint Mathematics Meetings. San Antonio, TX: Mathematical Association of America.

Triola, M. J. (2004). *Elementary Statistics* (9th Edition). Boston: Addison-Wesley.