

Conveying the Core Concepts
John D. McKenzie, Jr., Babson College
John D. McKenzie, Jr., Babson College, Babson Park, MA 02457

KEY WORDS: activities; applets; computational software; conceptual software

Abstract

What are the most important statistical concepts that students should learn in their first applied statistics course? How, in the context of a traditional course, especially one positioned as a service course, can we assure that students learn, understand, and internalize these concepts so that they can apply them effectively elsewhere? The panelists will propose a list of essential concepts and discuss incorporating such concepts into an introductory course. They will illustrate three basic approaches to presenting them: manipulative activities, computational software, and conceptual software. After discussion among the panelists, the audience will be invited to share comments and insights.

What is a Core Concept in Statistics?

Before addressing possible core concepts in statistics, let begin by proposing an operational definition of a core concept. For this paper it is a big idea or fundamental principle. It is not a specific application, method, or technique. Nor is it a skill, such as the use of software, or an attitude or value, such as the need for today's citizens to understand statistics.

A core concept in statistics is definitely not defined by one of the following explanations that were discovered in a literature search: a statistical characteristic of a time series or a kind of mathematical concept! Still such explanations are not that unusual according to Batanero (2001): "Notwithstanding its axiomatic foundation, Statistics may well be the only branch of Mathematics where debate is still going on about the interpretation of basic concepts".

Why is the Identification of Core Concepts Important?

In order to identify a list of possible core concepts in statistics, the following sources were examined: recent conceptual and applied statistical textbooks; ASA, MAA, AMATYC, and NCTM guidelines; MAA Notes; the Advanced Placement (AP) Statistics

course description, and numerous articles and presentations.

This literature led to widely varying views on the number of core concepts in statistics. One 2004 text stated in its preface that there was only one core concept in statistics: sampling distributions. The AP Statistics course identifies four major conceptual themes: exploring data: describing patterns and departures from patterns; sampling and experimentation: planning and conducting a study; anticipating patterns: exploring random phenomena using probability and simulation; and statistical inference: estimating population parameters and testing hypotheses. Another 2004 text listed 225 key concepts.

Some of this confusion may arise from the presence of at least three types of introductory statistics courses. The first is the statistical literacy/reasoning/thinking course. It emphasizes concepts. The second is applied statistics course. Often a service course, it stresses applications. Third, there is the mathematical statistics course, with its emphasis on theory.

But for over a decade there have been calls for incorporating statistical concepts in all of these introductory courses. As one example, Dick Scheaffer in 1992 stated that "The introductory course in statistics should focus on a few broad concepts and principles, not a series of techniques." As another example, the proposed Guidelines for Teaching Statistics: College, which were presented at a JSM session the next day, recommends that these courses should "Stress conceptual understanding rather than mere knowledge of procedures". Most of these calls have suggested that fewer topics are better for most courses ("Less is more!").

Thus there is a pressing need for a list of core concepts so that all instructors, but especially new instructors, can address related concerns. As two examples, how to avoid underestimating the difficulty students have in understanding these concepts and overestimating the understanding of these concepts by students. In addition, it is important to have a list of such concepts so instructors know how to incorporate them into introductory courses, especially the applied and mathematical statistics courses.

Identifying Core Concepts in Statistics

From a perusal of the literature the author constructed a list of 30 possible core concepts in statistics. This list was distributed to the members of the panel's audience in a Core Concept Survey and asked to mark all appropriate core concepts in the second column, the three most important core concepts in the third column, and the three most difficult core concepts in the fourth column. The audience was also encouraged to identify other possible core concepts at bottom of survey. And, while the results of the 56 submitted surveys were tabulated, the author's fellow panelists (Jerry Moreno, Robert H. Carver, Chris Lacke, and Ron Weiers) showed how they attempt to teach core concepts in statistics using software and activities. One of the concepts that each illustrated was variability because the panel organizers felt that this was the most important core concept.

Results of the Core Concept Survey

The results of the Core Concept Survey are present in the Appendix on the next page. The second column indicates the number of responses that considered each concept a core concept. The six concepts

marked most frequently were variability (54 marks), association vs. causation (46), randomness (43), significance (statistical vs. practical) (43), data collection methods (experiments, observations, surveys) (43), and sampling distributions (law of large numbers, CLT (central limit theorem)) (40). Only four concepts were infrequently marked: transformations (1), process (4), cross-sectional vs. longitudinal data (6), and the regression effect (6). The rest of the concepts received between 18 and 36 marks.

The third column indicates the number of responses that considered each concept a top 3 core concept. Variability with 38 marks was definitely one of these concepts. Other possible most important core concepts were association vs. causation (16), sampling distributions (law of large numbers, CLT) (13), data collection methods (experiments, observations, surveys) (12), and hypothesis test (critical values, p-values, power) (11).

Only two concepts appear to be consensus most difficult core concepts. They are hypothesis test (critical values, p-values, power) and sampling distributions (law of large numbers, CLT) with 33 marks.

Appendix: Results of the Core Concept Survey

	Core Concept?	Top 3 in Importance?	Top 3 in Difficulty?
Association vs. causation	46	16	3
Assumptions	31	4	10
Bias	27	1	1
Center	33	3	0
Comparisons	21	1	1
Confidence interval	35	6	8
Correlation	27	1	3
Cross-sectional vs. longitudinal data	6	0	0
Data collection methods (experiments, observations, surveys)	42	12	2
Data exploration	24	4	0
Distributions	29	5	7
Graphing	30	5	0
Hypothesis test (critical values, p-values, power)	36	11	33
Independence	28	2	8
Least-squares regression	22	1	4
Models	21	2	6
Outliers (aspects of robustness)	18	0	0
Prediction	19	1	1
Proportion	23	0	0
Process	4	0	0
Qualitative (categorical) vs. quantitative (measurement) data	34	4	2
Randomness	43	7	4
Random sampling	35	5	2
Regression effect	6	0	2
Sampling distributions (law of large numbers, CLT)	40	13	33
Shape	25	0	0
Significance (statistical vs. practical)	43	7	8
Transformations	1	0	5
Uncertainty	30	5	1
Variability	54	38	6