Capstone assessment for the undergraduate statistics major

Matthew Beckman
Penn State University

CAUSE Webinar

September 10, 2019

Collaborators

- Matt Beckman (Penn State)
- Beth Chance (Cal Poly–San Luis Obispo)
- Kirsten Eilertson (Penn State)
- Jennifer Kaplan (Georgia)
- Kari Lock Morgan (Penn State)
- Paul Roback (Saint Olaf College)

Funding: Penn State Center of Excellence in Science Education

Assessment Outline

- **Goal: Measure learning outcomes of students upon completion of undergraduate statistics program (e.g. major)**
  - snapshot of student learning outcomes
  - facilitate cohort comparisons for program evaluation
  - comprehensive scope
- Constraints
  - closely aligned to (2014) ASA Curriculum Guidelines[1]
  - ease of use across institutions, instructors, years
- Outcomes
  - two capstone assessment tools (i.e., test & survey)
  - multi-year pilot
  - good psychometric quality

---

[1] American Statistical Association Undergraduate Guidelines Workgroup (2014). 2014 Curriculum guidelines for undergraduate programs in statistical science. Alexandria, VA: American Statistical Association. http://www.amstat.org/education/curriculumguidelines.cfm

(2014) ASA Guidelines for Undergraduate Programs in Statistical Sciences

data. Such skills underpin strategies for assessing and ensuring data quality as part of data preparation and are a necessary precursor to many analyses[43].

- Use of one or more professional statistical software environments[44]
- Data management using software in a well-documented and reproducible way[45], data processing in different formats, and methods for addressing missing data
- Basic programming concepts (e.g., breaking a problem into modular pieces, algorithmic thinking[46], structured programming[47], debugging, and efficiency)
- Computationally intensive statistical methods (e.g., iterative methods, optimization, resampling, and simulation/Monte Carlo methods)[48]
- Use of multiple data tools[49], so graduates are not wedded to one and are better able to learn new technologies[50]

Mathematical Foundations

The study of mathematics lays the foundation for statistical theory. Undergraduate statistics majors should have a firm understanding of why and when statistical methods work. They should be able to communicate in the language of mathematics and explain the interplay between mathematical derivations and statistical applications.

- Calculus (e.g., integration and differentiation)[51]
- Linear algebra (e.g., matrix manipulations, linear transformations, projections in Euclidean space, eigenvalues/eigenvectors, and matrix decompositions)
- Probability (e.g., properties of univariate and multivariate random variables, discrete and continuous distributions)[52]
- Emphasis on connections between concepts in these mathematical foundations courses and their applications in statistics[53]

Statistical Practice

Strong communication skills complement technical knowledge and are particularly necessary for statisticians; graduates need technical skills to perform analyses and communication skills to understand clients' needs and then effectively discuss results and conclusions. Important practical skills include the following:

43 See Zhu et al. (2013) "Data acquisition and pre-processing in studies on humans: What is not taught in statistics classes?" The American Statistician, 67(4):235–241, which includes a series of skills: (1) get to know the study; (2) assess the validity of variable coding; (3) assess data entry accuracy; (4) perform data cleaning; and (5) edit identified data errors.

44 Although we acknowledge that Microsoft Excel is a common platform for data exchange, we do not recommend it as a primary analysis environment.

45 Appropriate environments could include R, Python, and SAS, complemented by tools including shell scripts and knitr.

46 Futschek (2006) defines algorithmic thinking as a set of abilities related to constructing and understanding algorithms: (1) the ability to analyze a given problem; (2) the ability to precisely specify a problem; (3) the ability to find the basic actions that are adequate to the given problem; (4) the ability to construct a correct algorithm to a given problem using basic actions; (5) the ability to think about all possible special and normal cases of a problem; and (6) the ability to improve the efficiency of an algorithm. Futschek, G. (2006). "Algorithmic thinking: The key for understanding computer science." In R. Mittermeir (Ed.), Informatics Education—The Bridge between Using and Understanding Computers (Vol. 4226, pp. 159–168). Berlin/Heidelberg: Springer. We consider this to be a necessary, but not sufficient component of "computational thinking."

47 We define structured programming as the ability to use functions and control structures (e.g., "for" loops).

48 This recommendation is consistent with the efforts of Conrad Wolfram and the Computer-Based Math initiative, www.computerbasedmath.org and www.ted.com/talks/conrad_wolfram. The incorporation of these tools may be particularly valuable at the bachelor's level, since students will generally have less technical knowledge (and need to be able to simulate to generate insights and/or check analytic results).

49 Students should develop the capacity to manipulate formats such as CSV, JSON (JavaScript Object Notation, a data interchange format that is easy to read, parse, and generate; see Nolan and Temple Lang (2014) XML and Web Technologies for Data Sciences with R), XML, databases (see, for example, Ripley (2001) "Using databases with R," R News, 1(1):18–20 and Wickham (2011) "ASA 2009 Data Expo," Journal of Computational and Graphical Statistics, 20(2):281–283), and text data. Because many faculty were not trained in these technologies, continuing education in this area needs to be made a priority.

50 We are not prescriptive regarding which technologies are incorporated into the curriculum, as long as they are sufficiently flexible and powerful. Many undergraduate statistics students develop expertise in environments such as R/RStudio, Python, and SAS.

51 Multivariate calculus is recommended.

52 Markov chains are a useful topic for undergraduate majors in statistics.

53 This linkage includes topics such as the delta method. In addition, many students might benefit from exposure to modeling and simulation in their mathematics courses as a way to reinforce their computational skills.

Comprehensive Undergraduate Statistics Program (CUSP) Assessment Strategy

- Test Blueprint (Link)
- 95 competencies cited in 2014 ASA Guidelines
- single assessment tool likely not sufficient

# Competencies	ASA Guidelines Topic
37	Statistical Methods & Theory
16	Data Wrangling, Computing, & Data Science
11	Mathematical Foundations
18	Statistical Practice
9	Problem Solving
4	Discipline-Specific Knowledge

Comprehensive Undergraduate Statistics Program (CUSP) Assessments

- Indirect assessment–CUSP Survey
  - inventory of all 95 competancies cited in ASA Guidelines
  - survey data self-reported by students
  - approx. 10-15 minutes duration
  - several cohorts from single institution
- Direct assessment–CUSP Test
  - selected response test
  - approx. 1 hour duration
  - multiple institutions w. single cohort
  - single institution w. multiple cohorts

Indirect assessment–CUSP Survey

- **Benefits**
  - Easy implementation
  - Can be administered multiple times if desired
  - No problem including topics we don't teach
  - Includes demographics
- **Risks/Issues**
  - Reliability of self-reporting
  - Over/Underconfidence with content exposure
- Indirect Assessment (Survey Link)

Excerpt

Statistical Theory

(scale: [1] very low / never learned; [2] low; [3] fair; [4] good; [5] very good; [6] excellent; [7] exceptional)

Please rate your **current level of knowledge/competency** related to:

	1	2	3	4	5	6	7
Distributions of random variables	○	○	○	○	○	○	○

Example Use

- Indirect assessment tool (i.e., Survey) administered at key program milestones
  - first-year seminar
  - midpoint course(s)–if possible
  - beginning & end of capstone course
- informative for annual program evaluation data

Direct assessment–CUSP Test

- Selected response assessment tool with broad coverage
- 33 tasks; some with multiple parts
  - 9 testlets
  - 24 conventional MC questions
- several tasks/subtasks assess multiple competancies
  - score adjustment for successive competancies (1, 1/2, 1/4, . . .)
  - 86 points possible
- some tasks adpted from other instruments (with permission)
  - 2 from the REGRESS assessment[2]
  - 9 from the CAOS assessment[3]

---

[2] Enders, F. (2013). Do clinical and translational science graduate students understand linear regression? Development and early validation of the REGRESS quiz. *Clinical and Translational Science, 6*(6). p. 444-451.

[3] delMas, R., Garfield, J., Ooms, A., Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*. p. 28-58.

CUSP Test

- Instructor Preview (link)
    - **preview is not for classroom use**
    - password protected

Excerpt (partial item)

driver or passenger side.

Study design dictates appropriate statistical analysis, but often there is more than one reasonable approach to the analysis. Evaluate whether each of the following analysis suggestions is VALID or NOT VALID for testing and estimating the difference in durability for the two brake pad materials:

	Valid	NOT Valid
paired t-test for brake pad difference of each car (DriverSide - PassengerSide)	○	○
paired t-test for brake pad difference of each car (Experimental - Standard)	○	○
ANOVA with car as a blocking variable	○	○

CUSP Test

- **Benefits**
  - test statistical “reflexes” of students
  - built-in “CAOS” subtest
  - objective measure of student competancies
    - for individual students
    - for a cohort of students
    - aggregate useful for program evaluation
  - Easy implementation
- **Risks/Issues**
  - Variable use conditions may jeopardize comparisons
  - Scope constrained by test fatigue
  - Includes topics we don’t necessarily teach
  - Longer to implement

Example Use Cases

- Penn State
  - Indirect assessment tool (i.e., Survey) administered pre & post (in addition to first-year students)
  - Direct assessment tool (i.e., test) positioned as a midterm in the capstone course
  - benchmarking student skills and competancies against ASA Guidelines
  - allows several weeks to address areas of need prior to graduation
  - useful for annual program evaluation data
- Other Institutions
  - no course credit
  - homework, extra credit, etc
  - typically open notes

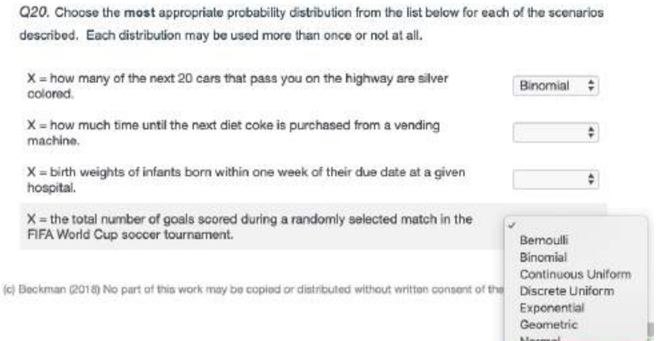
Preliminary Item Functioning Analysis

- Benchmarks for item statistics[4]
  - Unidimensionality assumed by common methods of assessment evaluation
  - Cronbach alpha (reliability)
  - descrimination > 0.15 preferred
  - $0.6 <$ proportion correct < 0.9
- Results
  - PCA evidence supports unidimensionality
  - Cronbach alpha $= 0.81$
  - $30/33$ items with discrimination > 0.15
  - $9/33$ items in recommended difficulty range
  - $21/33$ items with $> 50\%$ correct

[4] Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge: New York.

Item discrimination

- Item discrimination < 0.15
  - (21% correct) Validity of models aligned to a study design
  - (3.6% correct) Strategies to maximize likelihood
  - (40% correct) CAOS task about CI interpretation
- Best item discrimination
  - (discrim $= 0.59$) Probability distributions task
  - (discrim $= 0.50$) Histograms & std deviation task
  - (discrim $= 0.46$) OLS regression assumptions task


Q20. Choose the **most** appropriate probability distribution from the list below for each of the scenarios described. Each distribution may be used more than once or not at all.

X = how many of the next 20 cars that pass you on the highway are silver colored. — Binomial

X = how much time until the next diet coke is purchased from a vending machine.

X = birth weights of infants born within one week of their due date at a given hospital.

X = the total number of goals scored during a randomly selected match in the FIFA World Cup soccer tournament.

Bernoulli
Binomial
Continuous Uniform
Discrete Uniform
Exponential
Geometric

(c) Beckman (2018) No part of this work may be copied or distributed without written consent of the

Future work

Shorter term goals

- Streamline logistics for wider implementation
- Link CUSP Survey data to CUSP Test outcomes
- Expand item bank

Longer term goals

- Experimentation with short/long forms
- Alternative or additional tools for more complete alignment to ASA Guidelines

Acknowledgments

- Advisory input
  - Nick Horton
  - Allan Rossman
- Pilot testers
  - Heather Smith
  - Andrew Schaffner
  - Nicole Lazar
  - Lynne Seymour
  - Paul Roback
  - Kirsten Eilertson
  - Dave Hunter
  - Christian Schmid
  - Daisy Philtron

Discussion

Backup slides

CUSP Test blueprint alignment to ASA Guidelines

Section	Subsection	Target Weight (%)
Statistical Methods and Theory	Statistical Theory	18.0
Statistical Methods and Theory	Exploratory Data Analysis	6.0
Statistical Methods and Theory	Design of Studies	18.0
Statistical Methods and Theory	Statistical Models	18.0
Data Wrangling Computation and Data Science	Software and Tools	0.0
Data Wrangling Computation and Data Science	Accessing and Wrangling Data	4.5
Data Wrangling Computation and Data Science	Basic Programming Concepts	1.5
Data Wrangling Computation and Data Science	Computationally Intensive Statistical Methods	4.0
Mathematical Foundations	Calculus	0.0
Mathematical Foundations	Linear Algebra	0.0
Mathematical Foundations	Probability	2.5
Mathematical Foundations	Connecting mathematical foundations & applications in statistics	2.5
Statistical Practice	Communication	0.0
Statistical Practice	Collaboration	0.0
Statistical Practice	Ethical Issues	5.0
Statistical Practice	Opportunities for Authentic Practice	0.0
Problem Solving	Complex open-ended problems	2.2
Problem Solving	Scientific method and statistical problem-solving cycle	12.8
Discipline-Specific Knowledge	Discipline-Specific Knowledge	5.0

Scree plot of CUSP test data