Cognitive Challenges of Scientific Inference

Scientists must incorporate both the reliability of evidence and the scope of inference, yet both are complex phenomena that can be difficult to understand.

Difficulties with reliability of evidence
- Tversky & Kahneman (1973) found quantitative psychologists vastly overestimated replicability of findings from small studies.
- Underpowered studies lead to many “voodoo correlations” (Fiedler, 2011) being reported in the literature.
- Adults frequently neglect sample size (see Linn et al., 2011), which statistical training reduces but does not eliminate (e.g., Tong, Krantz, & Nosib, 1986).

Difficulties with scope of inference
- Introductory statistics students struggle to understand when study design warrants generalizing to a population or concluding causation (e.g., Derry et al., 2000).
- News media reporting routinely overgeneralizes and adds inappropriate causal attributions (Schield, 2010).
- Scientists face many cognitive biases (Nuzzo, 2015) and overstate their own findings (see Utts, 2003).

Statistics education research can contribute to metascience’s investigation of how scientists interpret and treat evidence:
- Students face unique challenges in learning statistics that require specialized pedagogical research (Garfield & Ben-Dr, 2007).
- Statistics education research provides perspectives on students’ understanding at all levels, from kindergarten to the workforce.
- Our studies of introductory college statistics students’ understanding can illuminate the challenges in conceptualizing and adequately addressing threats to internal and external validity.

Teaching Inference with Simulation

Statistics educators increasingly use simulation and resampling techniques to teach frequentist inference (Tinto et al., 2015):
- Simulation, bootstrap resampling, and randomization may help students focus on the unified logic of the sampling distribution (casists, 2011).
- Displaying uncertainty via concrete realizations of a stochastic process may map on to human cognitive faculties more naturally than abstract distributions (Jagermeer & Mihof, 2010; Nelsen et al., 2013).
- Simulation-based introductory statistics curricula have some evidence of providing advantages in student understanding and outcomes (Tinto et al., 2012).

Example simulation-based activity (from Zieffler et al., 2017)
In Hamline & Bloom (2007), 14 out of the 16 (87.5%) babies chose the “Helper” toy. Does this study provide evidence that infants notice and prefer the “Helper”?

1. Students set up null model visually in TinkerPlots e.g., a spinner with 50% helper and 50% hinderer, or could be two balls “helper” and “hinderer” chosen with replacement.
2. Students plot the result of a single trial under null hypothesis.
3. Students collect the percentage of means of many trials, and evaluate how rare results at or above 87.5% would be under the null model.

Study 1: Reliability of Evidence (Brown, 2019)

In one year, which hospital do you expect to have more days with more than 60% boys?
- The larger hospital
- The smaller hospital
- About the same (rounded to whole value)

As sample size increases, we grow more certain that a statistic is close to the population value. This fundamental principle is called the Empirical Law of Large Numbers (Freedman, 1972).

Common confusions include:
- Ignoring sample size, believing that similarity to population is only factor
- Confusing frequency and sampling distributions
- Attending more to the ratio of the sample to the population size

Prior interventions:
- Demonstrate that sample size is responsive to training
- Rely on learning a rule (e.g., Mann, :1966) and/or inferring a principle from demonstrations
- Do not support students in exploring the mechanisms of power and precision

Results
- Nearly all students successfully used swapping reasoning, but application depended on context and student experience.
- Students were able to recognize that having more variability, or a mean closer to hypothesized value, meant a larger sample size was needed.
- Students had difficulty interpreting many samples and often argued based on results from a single sample.
- Students successfully drew on their everyday experiences (getting an A at the end of a course) as opposed to beginning to explain Empirical Law of Large Numbers

Limitations
- Sample size of 1!
- Mechanistic coding had 1 coder
- Attribution to activities unclear
- Low ecological validity
- Not classroom-scalable

Study 2: Scope of Inference (Fry, 2017)

According to statistics education recommendations (e.g., Utts, 2003), students should understand:
- Random sampling tends to produce representative samples, allowing for generalization to a population.
- Random assignment tends to balance out confounding variables between groups, helping to enable cause-and-effect conclusions.

Students have difficulties such as:
- Confusion between random sampling and random assignment (Derry et al., 2000)
- Disbelief that random assignment can help enable causal claims (Sawilowsky, 2004)
- Believing that larger samples are always better than smaller samples, regardless of method (Wagler & Wagler, 2013)

Overall, evidence of gains in learning goals related to study design and conclusions, especially those related to understanding:
- The purpose of random assignment
- Correlation does not imply causation
- A small, but noticeable portion of students experience difficulties such as:
- Confusion between random sampling and random assignment
- Giving sample size more importance than sampling method

Sample item showing some evidence of lingering confusion between random sampling and random assignment:

Researchers conducted a survey of 1,000 randomly selected adults in the United States and found a strong, positive, statistically significant correlation between income and the number of containers the adults reported recycling in a typical week. Can the researchers conclude that higher income causes more recycling among U.S. adults? Select the best answer from the following options.

Response
- Yes, the sample was randomly selected, so correlation can be inferred.
- No, the lack of random assignment does not allow causation to be inferred.
- Yes, for statistically significant result, a causal factor must be inferred.
- No, the sample was not randomly selected, so correlation can be inferred.

Future Directions in Teaching Scientific and Statistical Inference

Statistical Understanding and the Reproducibility Crisis
- Better instruction is needed to support scientists and non-scientists in statistical reasoning. Simulation-based representation may be a useful tool.
- Much more research is needed into understanding how these crucial types of reasoning can be improved.
- Another crucial topic is people’s understanding of the implications of multiple testing and selective inference (e.g., Benjamini & Hochberg, 1995), another major contributor to the reproducibility crisis.

Building more experience with statistical processes
- These complex concepts may need more experience beyond standard statistical courses
- Earlier, deeper, and wider accessible introductions to statistical reasoning can start in early primary school (Franklin, 2005)
- Reliability and scope of inference can be integrated into science education to give students more experience
- Arithmetic has moved from an elite to a basic literacy skill. It may be time for the same transition for reasoning with data and uncertainty

References
For a complete list of references, please see the original publication.