Think-aloud interviews and cognitive task analysis to identify misconceptions in undergraduate statistics

Mikaela Meyer, Josue Orellana*, and Alex Reinhart

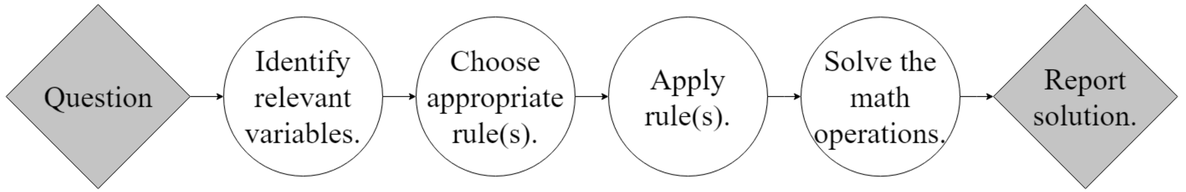
Dept. of Statistics & Data Science, *Machine Learning Dept. and Center for the Neural Basis of Cognition, Carnegie Mellon University

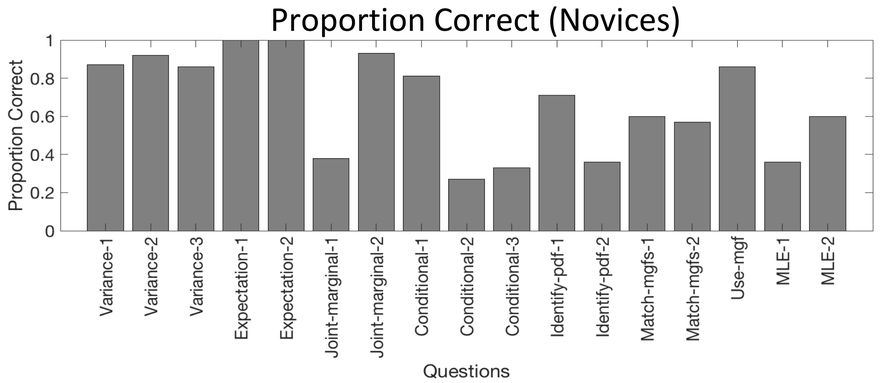
Cognitive Science Techniques

- **Think-Aloud Interviews:** Respondents are asked to say everything they are thinking about while answering a question [1]
- **Cognitive Task Analysis (CTA):** Create an outline of the steps required to solve a problem
- Think aloud interviews can detect specific misconceptions as departures from this prescribed cognitive task model.

Methods

- Created 25 questions on introductory statistical inference topics. Examples:
  - Find f(y|x), given f(x,y), f(x) and f(y)
  - If Z = 3X + 2c, where X ~ N(0,1) and c is a constant, find Var(Z)
  - Ensured notation matched the notation used in CMU's introduction to probability theory undergraduate course
- Conducted eight think-aloud interviews with "experts" (Ph.D. students) and sixteen with "novices" (undergraduate students)
  - 60 minute interviews
  - Participants paid $20
  - Audio from interviews recorded





What we heard in think-aloud interviews

The log-likelihood for $x_1, x_2, \ldots, x_n$ i.i.d. samples from a univariate normal distribution is:

$$\log \mathscr{L}(\mu, \sigma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Find $\hat{\mu}_{\mathrm{MLE}}$.

Experts:

- "So it's gonna be the mean, but let's prove it."
- "And just to check that is a maximum, you take the second derivative and check that it is hmm check that it is negative, so that it is a maximum"

Novices:

- "I always get weirded out when I have to do the derivative of a sum, like I don't really know if there's rules…"
- "So we just take the derivative of this with respect to... what do you call it, sigma, right? Yeah, yeah, so sigma. Or is [it] with respect to sigma, or with respect to mu?"

Next Steps

- Analyze remaining questions.
- Conduct more think-aloud interviews with novices.
- Work with instructors to develop improved teaching strategies.
  * Based on what we've learned thus far, we feel that identifying relevant variables is a difficult step for students.
  * We also have realized through these interviews that some students need more practice with some calculus skills.

References

1. Ericsson, K. A. and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data.* MIT Press, 2nd edition

Acknowledgements

Carnegie Mellon University's GSA/Provost GuSH Grant funding was used to support this project. Thanks to the Department of Statistics & Data Science at CMU for their financial support and to all the students who participated in interviews.