

Plenary Talk in Statistics

What are the limits of inferences from lousy data, and why should we care?

Paul Gustafson, University of British Columbia

In health research and other fields, the observational data available to researchers often fall short of the data that would ideally be available, due to inherent limitations of study design and data acquisition. Were they available, the ideal data might readily be analyzed via straightforward statistical models with desirable properties such as parameter identifiability. This gives the “textbook” situation of estimators which tend to the right answer as the sample size grows, at a rate given by the square-root of the sample size. This talk surveys the “non-textbook” (and very real) situation where realistic statistical models for the available data won’t be identified. We consider Bayesian parameter estimation in this setting, focussing on how much uncertainty would remain about a quantity of interest if an infinite dataset was observed. Surprising answers ensue in some situations. Implications for questions of whether a given study should be undertaken, and how much data should be collected, will be discussed.