

Judgment, AI, and Statistics

Fred Philips

College of Engineering & Applied Sciences, and
Alan Alda Center for Communicating Science, SUNY-Stony Brook, and
TANDO Institute, USA (fred.phillips@stonybrook.edu)

First received 27 February 2025. Published online 25 June 2025.

Abstract

An experienced journal editor remarks on an apparent imbalance in preparing young management researchers to do empirical work. Focusing on the nature of statistical analysis, the cultural habits of various management disciplines, and touching on the research use of artificial intelligence, this essay brings these elements (and other aspects of the research process) together to suggest that judgment is a critical element of successful empirical research. It urges re-examination of the graduate Research Skills syllabus.

DOI: 10.62478/ITBA9624

IPNS: k2k4r8o0qjwvqpb8nz5brp55oiot3k6xai9zu0rn9srrrrfijmfat284

KEY WORDS: Statistics; Education; Research; Science

1. The problem

Lately I've seen several papers in which the sophistication of the statistical analysis overshadows the overall study logic. They reinforce my longstanding impression that we senior faculty have erred in what we emphasize when we train young researchers.

This mis-emphasis arises in part because it's easy to teach statistical software, harder to convey the epistemological basis of statistical inference, and still harder to teach good judgment in study design and interpretation. As a result, we have allowed "Who's good at statistical math?" to underpin the status structure of the grad students in our labs. New graduates carry this skewed view into their professional careers.

2. Toward a new balance

With apologies for belaboring the basics, here's how we can re-balance education in data-based management research. The answers involve clearer exposition of the meaning and limitations of statistics, and of the role of judgment in the more complex analyses (the ever-popular structural equation modeling –SEM – in particular). It involves noting that science must examine the observer as well as the observed, as the traditions of our sub-disciplines tend to put blinders on their researchers. It involves more careful use of artificial intelligence (AI).

Statistics

First, emphasize to students that statistical inference doesn't prove anything. It can only lend a level of confidence in the results of an analysis. Analysis needs to walk hand in hand with a well-structured

research plan, sound data, and good judgment in interpretation. That combination is what makes a study convincing.

Statistical confidence, in turn, needs to be clarified. Rejecting (or failing to reject) a hypothesis at the 90% level means that if we were to repeat the experiment 100 times with 100 independent random samples, we would expect the same result approximately 90 times (Phillips 2016). Yet in a fast-changing business environment, we'll never be able to replicate an experiment many times under identical conditions. Discuss: Where does that leave us as management researchers?

This frequentist view of statistics thus seems fanciful, reasonable in principle but not in practice. Yet it's perfectly plausible compared to the alternative! Students may be exposed to the Bayesian view, but should not be encouraged to rely on it. Bayesianism – the most prominent alternative to the frequentist view – treats probability as a measure of belief or certainty about an event.

Belief? Certainty? It's absurd to think that simple equations could say anything meaningful about our inner psychology. Moreover, Bayesianism highlights the question of judgment. New graduate students can easily master the mathematics needed for either statistical paradigm. Their judgment, on the other hand, will mature as their studies progress and as their careers progress. They should not be encouraged to use subjective methods early on.

Even simple factor analysis requires judgment. (How many factors?) A structural equation model presents many needs for judgment in reducing variable to factors. The point here is that students want to see their research published. Even in double-blind review situations, where the age of the researcher is not evident, the consequences of the researcher's judgments are evident. It's risky to submit a judgment-based paper early in one's career, less risky for a more mature researcher.

Now what about "independent random samples"? Many times I've asked a first-year graduate class what 'random sample' means, without getting an accurate answer. (Same for 'independent.')

Colleagues, please ensure the correct definition is burned into students' brains, along with the idea that a complicated analysis means nothing, and is not publishable, if the sample is not random: The mathematics of statistical inference depend utterly on the randomness of samples. Then tell them, "Don't even think of turning in an empirical paper that doesn't state how you obtained a truly random sample – or doesn't convincingly explain how your sample departs from randomness only negligibly."

A separate discussion is needed about randomness when the student draws the sample from a huge online (secondary) data set. And still another discussion on how to present and interpret simple analyses of nonrandom samples.

Students also tend to forget that samples are for making inferences about a population – and their reports must identify and delimit the population that's being studied!

Disciplinary tradition

I asked a Paper Development Workshop author, "Why did you use this technique?"

He replied, "This is the customary statistical technique in my subfield of management."

Well! He did not say it was the right technique for the research question and for the data set – he said it was the customary technique. Meaning, it was what he was taught, what he habitually turns to, and what he believes editors of journals in his field are looking for and comfortable with. And often, one confidently imagines, a round peg forced into a square hole.

As Thomas Kuhn (2012) would say, science is a social endeavor. Scientific results are functions of the researcher[1] and the researcher's community, as well as of the object of study. Yet innovative science means breaking free of consensus science's mold.

This can be accomplished by teaching a variety of statistical techniques[2], teaching how to match them with the research question and with available data formats, and awarding promotion credit for publications in good journals that are outside the academic department's official preferred set. Meaning, journals whose editors have more diverse but still rigorous views concerning methodologies.

Artificial intelligence

“Evaluating the output of [AI] models requires expertise and good judgment. . . . AI is likely to widen workforce divides”[3]. That is, in our context, the divide between researchers who exercise the required judgment and those who do not.

LLMs return the most frequently expressed views, rather than the maverick or leading-edge ideas. Again, a consensus trap. “Researching is thinking: noticing. . . gaps in the conventional wisdom. . . The risk of outsourcing all your research to a super-genius assistant is that you reduce the number of opportunities to have your best ideas” [4].

In other words, relying too much on AI is just like relying too much on statistical inference: Excessive dependency on either AI or statistics detracts from an innovative and well-rounded research project (Phillips 2024).

3. Summary

This essay has emphasized judgment as an essential element of research, but has not yet defined it. I consider it a combination of: Integrating one’s life experiences, including work, travel, family and so on; recognizing biases and fallacies; reading critically and widely (not just academic journals); conversing with diverse kinds of people, including academic colleagues and practicing managers, while respectfully entertaining contrary opinions; caring sincerely for the research topic; and consulting one’s own heart and gut as well as head.

The essay may appear to contradict itself, urging judgment as essential to an overall study, yet urging young researchers not to inject it too much into their statistical analyses. One definition of a gentleman is a man who knows how to play the saxophone but reliably refrains from doing so. In the same way, maturing researchers cultivate judgment but apply it in their analyses only judiciously at first, then more liberally as their research experience accumulates.

Of course graduate courses in research skills must teach statistical analysis. They should also include discussion of all the items mentioned in the present article, as well as how to craft a research question and a sampling plan; how to tell good data from questionable data; how to interpret results; how to write well; and how to cautiously assess results’ generalizability. In this way, the next generation of graduates will be not only good analysts, but good researchers.

Footnotes

1 We may put a fine point on Kuhn’s assertion by noting that Reverend Bayes was a Presbyterian minister, perhaps predisposed to favoring belief over evidence. A Wikipedia article, https://en.wikipedia.org/wiki/Thomas_Bayes, suggests Bayes’ interest in probability arose in the course of an argument with David Hume concerning whether to believe in miracles. Though the belief element persists to this day in Bayesian research circles, Bayes’ Theorem does have a useful, belief-free mathematical interpretation. It can be taught to beginning statistics students without reference to psychology or theology. 2 Not that the researcher will use them all. However, when it’s feasible to use multiple methodologies – “triangulation” – stronger papers can result. 3 *The Economist* Feb. 15, 2025, p.57. 4 *The Economist* Feb. 15, 2025, p.63.

References

- Kuhn, Thomas S. (2012 edition) *The Structure of Scientific Revolutions*, University of Chicago Press.
- Phillips, Fred (2016) “Deciphering p values: Defining significance.” *Science* 05 Aug: Vol. 353, Issue 6299, p.551.
- Phillips, Fred (2024) “From My Perspective: Artificial Intelligence and Research Productivity.” *Technological Forecasting & Social Change*, 212 (2025) 123937.