

# Statistical significance: moving to a world beyond $p < 0.05$

Anton H. Westveld

Research School of Finance, Actuarial Studies & Statistics, ANU

June 2019

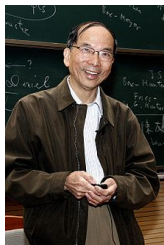


Australian  
National  
University

# What is Statistical Science?



- Statistics is the **science of learning from data**.



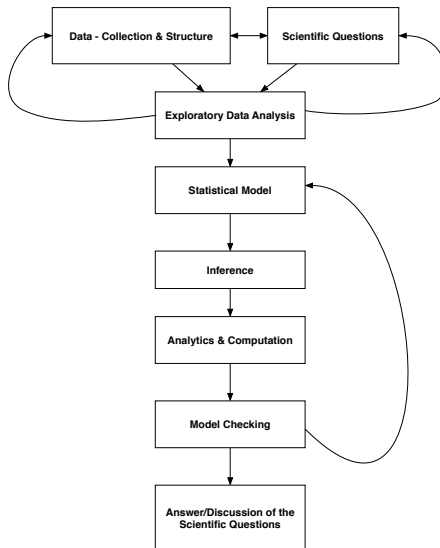
- Professor Jeff Wu in November 1997, gave a talk for his appointment to the H. C. Carver Professorship at the University of Michigan titled:

**Statistics = Data Science?**

<http://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf>

$$\begin{aligned} &f(\text{Data}) + f(\text{Mathematics}) + f(\text{Computation}) + f(\text{Subject Matter Knowledge}) \\ &= \text{Statistical Science} \end{aligned}$$

# Flow of Statistical Scientific Inquiry



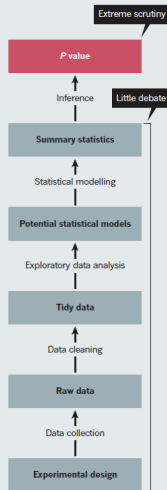
# Jeff Leek & Roger Peng's Take



Australian  
National  
University

## DATA PIPELINE

The design and analysis of a successful study has many stages, all of which need policing.



-Leek and Peng (2015)

**"Accept uncertainty. Be thoughtful, open, modest"**

-Wasserstein, Schirm and Lazar (2019)

- Prioritize effort for sound data production: the planning, design, and execution of the study. (Tong, 2019)
- Staff the study team with members who have the necessary knowledge, skills and experience – statistically, scientifically, and otherwise. (Brownstein et al., 2019)
- Understand that subjective judgments are needed in all stages of a study. (Brownstein et al., 2019)
- Make all judgments as carefully and rigorously as possible and document each decision and rationale for transparency and reproducibility. (Brownstein et al., 2019)
- Use flexible descriptive methodology, including disciplined data exploration, enlightened data display . . . for exploratory research. (Tong, 2019)
- Look for and present results from many models that fit the data well. (Lavine, 2019)

- Don't use bright dividing lines.

*We conclude, based on our review of the articles in this special issue and the broader literature, that it is time to stop using the term “statistically significant” entirely. Nor should variants such as “significantly different,” “ $p < 0.05$ ,” and “nonsignificant” survive, whether expressed in words, by asterisks in a table, or in some other way. -(Wasserstein, Schirm and Lazar, 2019)*

- Report the outcome of studies as effect sizes summarized by confidence intervals (CIs) along with their point estimates. (Mathews, 2019)
- Make full use of the point estimate and width and location of the CI relative to the null effect line when interpreting findings. (Mathews, 2019)
- Interpret interval estimates as “compatibility intervals,” showing effect sizes most compatible with the data, under the model used to compute the interval; do not focus on whether such intervals include or exclude zero. [Amrhein, V., Trafimow, D., and Greenland, S.]



- Present the probability assertions about the magnitude of a treatment effects [and or hypotheses]. (Ruberg et al., 2019)
- Incorporate prior data and available information formally into the analysis . . . . (Ruberg et al., 2019)
- Employ quantitative utility functions to reflect key considerations from all stakeholders for optimal decisions via a probability-based evaluation of the treatment effects. (Ruberg et al., 2019)



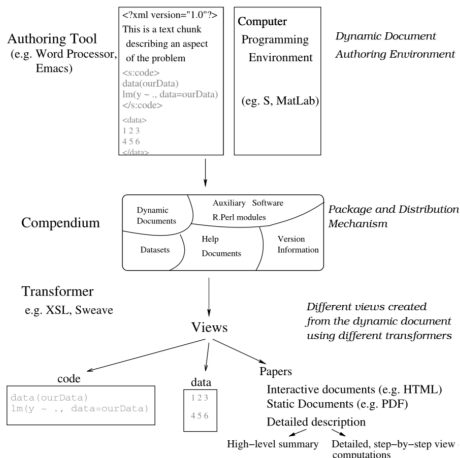
- Predict observable events or quantities that you care about. (Billheimer, 2019)
- Quantify the uncertainty of your predictions. (Billheimer, 2019)
- Consider the predictive distribution of a “treatment” effect for a single individual or a group.

- Replication: “re-performing the experiment and collecting new data.” (Peng, 2015; Patil, Peng and Leek, 2016)
- Two major components to a reproducible study:
  - “that the raw data from the experiment are available;”
  - “that the statistical code and documentation to reproduce the analysis are also available.”

# Replication & Reproducibility



- This call has been for some time: Gentleman and Lang (2007)
  - Tools are coming onboard (Sweave, R markdown)
  - Careful of Excel as a lab notebook - Dean Billheimer



p-value =



Australian  
National  
University

world o' el futuro

- Written on a painting at the Bio5 Institute at the University of Arizona

- Billheimer, Dean. 2019. "Predictive Inference and Scientific Reproducibility." *The American Statistician* 73.
- Brownstein, N., T. Louis, A. O'Hagan and J. Pendergast. 2019. "The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision-Making." *The American Statistician* 73.
- Gentleman, Robert and Duncan Temple Lang. 2007. "Statistical Analyses and Reproducible Research." *Journal of Computational and Graphical Statistics* 16.
- Lavine, M. 2019. "Frequentist, Bayes, or Other?" *The American Statistician* 73.
- Leek, Jeffrey T and Roger D. Peng. 2015. "P values are just the tip of the iceberg." *Nature* 529.
- Mathews, R. 2019. "Moving Toward the Post  $p < 0.05$  Era Via the Analysis of Credibility." *The American Statistician* 73.
- Patil, Prasad, Roger D. Peng and Jeffrey T. Leek. 2016. A statistical definition for reproducibility and replicability. Technical report bioRxiv: <https://doi.org/10.1101/066803>.

- Peng, Roger D. 2015. "The reproducibility crisis in science: A statistical counterattack." *Significance* .
- Ruberg, S., F. Harrell, M. Gamalo-Siebers, L. L. LaVange, J. Lee, K. Price and C. Peck. 2019. "Inference and Decision-Making for 21st Century Drug Development and Approval." *The American Statistician* 73.
- Tong, C. 2019. "Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science." *The American Statistician* 73.
- Wasserstein, Ronald L., Allen L. Schirm and Nicole A. Lazar. 2019. "Moving to a World Beyond " $p < 0.05$ "." *The American Statistician* 73.