

# INTEGRATIVE STATISTICS

Customized Statistical Analysis  
and Survey Research

11<sup>th</sup> Annual Public Performance Conference  
Suffolk University, Oct. 2018

## Pitfalls of Public Performance Analysis

Roland B. Stark  
Statistician and Research Consultant

### Introduction

Here I present lessons from my experience as a research and evaluation consultant, 2001-2018. I hope this informs your own research and aids your interpretation of research findings and that you pass on to colleagues some points that resonate with you.

### Pitfalls

#### 1. Confusing correlation with causation.

It takes extra effort to get to the bottom of cause-and-effect questions.

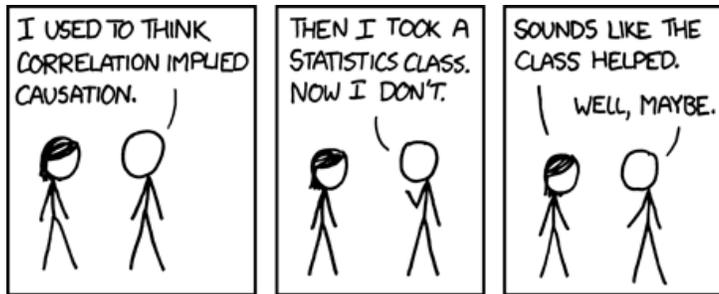
**STATISTICAL CONTROL** is no cure-all.<sup>1,2</sup>

Think through, again and again, your efforts to control for nuisance variables. Can you use random assignment? Complex sampling? Propensity scores? ... Have you queried colleagues extensively for additional ideas — without regard to rank?

---

<sup>1</sup> <http://www.reinforcedcare.com/resources/home-health-puzzle/>.

<sup>2</sup> My user-friendly intro to analyzing not just for correlation but for causation:  
<https://www.yellowbrickstats.com/documents/causation1.pdf>



from [www.XKCD.com](http://www.XKCD.com)

Also important: "Imply" can mean both "suggest" and "necessitate"!

Implications for practitioners and stakeholders: Using data analysis to draw only those causal conclusions that are justified is essential to sound evaluation. This section aims to use memorable examples to help participants make strides in that direction.

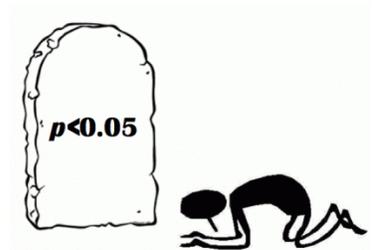
## 2. Over-emphasizing statistical significance and neglecting the size of an effect.

" $p < .05$ : **A** is significantly higher than **B**." OK! Is it 600% higher? 1% higher?

Produce confidence intervals – they're much more informative.<sup>3,4</sup>

Implications for practitioners and stakeholders: You'll understand how claims of statistical significance work. You'll be equipped to disregard a "statistically significant" finding when it's not practically significant. And you'll be able to interpret what confidence intervals say about the wider population.

from <http://emcrit.org/pulmcrit/demystifying-the-p-value/>

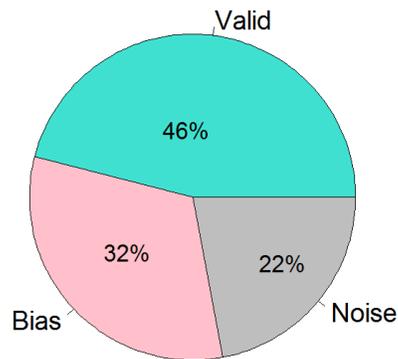


<sup>3</sup> The American Statistical Association has recently produced a position paper on this subject: <https://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108#.WzvXVFVKhCu>.

<sup>4</sup> For more on the insidious nature of significance testing, see <http://yellowbrickstats.com/insidious.htm>.

### 3. Waiting too long to consult a statistician.

- a. Validity and reliability of measures require advance planning. After all, the average survey question produces a mix of valid info, biased info, and random noise.<sup>5</sup>



Survey questions might require more options than Yes/No – or might need *fewer* options to be reliable or to avoid “splintering.”

Complex or “soft” constructs like “satisfaction” or “self-efficacy” (over-used!) at least need a scale carefully fashioned from multiple indicators.<sup>6</sup>

Survey grids invite bias in certain situations; you need to know which.<sup>7</sup>

Validation might call for “corroboration” using additional variables (*concurrent, predictive, or discriminant* validity).

---

<sup>5</sup> D. H. Doty and W. Glick in *Organizational Research Methods* (1998).

<sup>6</sup> A classic, very helpful guide is Spector, P. E. (1992), *Summated rating scale construction*. Newbury Park, California: State University Papers Series: Quantitative Applications in the Social Sciences.

<sup>7</sup> Roßmann, Gummer, and Silber in this month’s *Journal of Survey Statistics and Methodology*, <https://academic.oup.com/jssam/article/6/3/376/4349665>.

## Waiting too long to consult a statistician (continued)

b. **STATISTICAL POWER** may be..... too ..... low.

Power is the chance of finding an effect to be statistically significant (“conclusive”) – **if** the effect truly exists.

How frustrating, and wasteful, to be unable to say anything specific after an expensive study!

“All we can say is that Group A does better than Group B.”

“All we can say is that the correlation in the wider population is between .10 and .50.”

Power is based partly on sample size (and also on measurement precision, data variability, and your chosen threshold for significance).

Implications for practitioners and stakeholders: This section will deepen participants’ understanding of analysis of cause and effect; of validity and reliability; and of statistical power, a vital tool for preventing waste.

