PSY 503: Foundations of Psychological Methods

Lecture 2: Foundations? What? Why?

Robin Gomila

Princeton

September 2, 2020

Motivations

Why Are You Taking a Foundational Course About Statistics?

Foundational knowledge?

- In which domains of life have you acquired some foundational knowledge?
- When is foundational knowledge useful? Why spend time learning the foundations of a discipline or activity?
- When is it less relevant?

What methodological skills should psychological scientists develop?

- Two possible routes:
  - Simple algorithms: what to do and when?
  - Foundations: how and why we do what we do in specific circumstances?

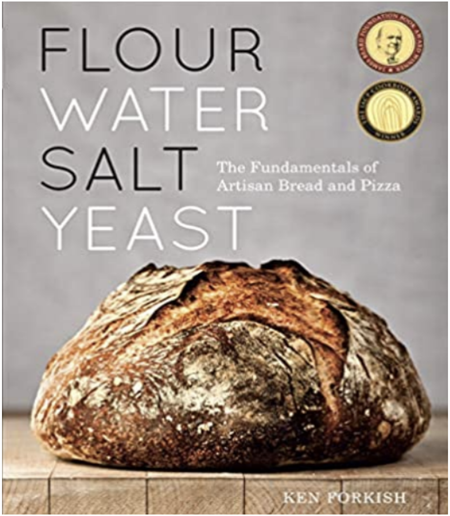
Making Bread: Simple Algorithms vs. Foundations

127 REVIEWS

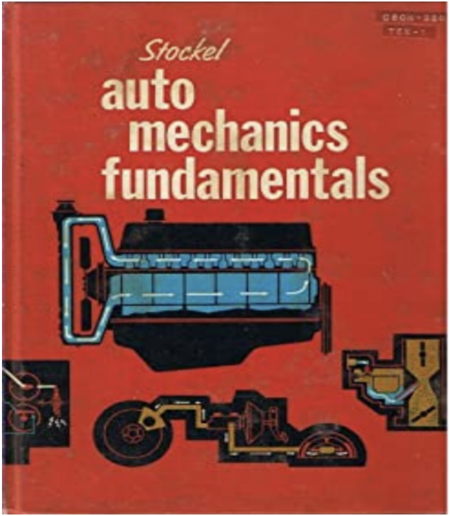
SHARES

With just five everyday ingredients, simple instructions, and no advanced baking techniques, this recipe for European-style crusty bread is a great introduction to yeast baking. It truly is "the easiest loaf of bread you'll ever bake" — thanks in large part to the high-protein of King Arthur Unbleached Bread Flour, which guarantees great texture and a high rise no matter how elementary a baker you may be!

PREP	BAKE	TOTAL
20 mins	20 to 25 mins	2 hrs 10 mins



Fixing Car: Simple Algorithms vs. Foundations



Norm Interventions: Simple Algorithms vs Foundations

26

The Construction of Social Norms and Standards

DALE T. MILLER
DEBORAH A. PRENTICE

Hamlet claimed, "There is nothing either good or bad, but thinking makes it so." Were Hamlet to have had a more social psychological turn of mind, he might well have said "There is nothing either good or bad but comparison makes it so." For the assessment of the self—its possessions, attributes, and accomplishments—is a largely comparative process. The idea that self-evaluation and self-experience are comparative dates back to the ancients (Hyman & Singer, 1968). The more recent contributions of social psychologists to the topic lie in their demonstrations of the scope and prevalence of comparative self-appraisal and in their delineation of its antecedents and consequences.

Of these two sets of contributions, the former is perhaps still the most significant. The social psychological literature contains numerous demonstrations of the dependence of our self-appraisals on comparison with others, across both judgments of abilities and attributes (e.g., Alicke, 1985; Campbell, 1986; Felson, 1986; Jones & Regan, 1974; Marsh & Parker, 1984; Morse & Gergen, 1970; Tabachanik, Crocker, & Alloy, 1983) and judgments of physical and material well-being (e.g., Bernstein & Crosby, 1980; Emmons & Diener, 1985; Frank, 1985; Lowenstein, Thompson, & Bazerman, 1989; Reis, Gerard, & Gibbons, 1993; Rice, McFarlin, & Bennett, 1989; Salovey & Rodin, 1984). In addition, there is now a substantial literature on the consequences of comparative judgment for self-evaluation and self-regulation (see Mischel, Cantor, & Feldman, Chapter 12, this volume; Tesser & Martin, Chapter 14, this volume). A sustained line of investigation has focused on when comparison processes result in assimilation of the self to a comparison standard and when they result in contrast between the self and the standard (e.g., when seeing an attractive person makes one feel more vs. less attractive). Recent models of social judgment have suggested a number of factors that determine when assimilation or contrast is more likely (see, e.g., L. Martin, 1986; Schwarz & Bless, 1992: Wegener & Petty, 1995). Researchers have also examined the psychological and behavioral consequences of comparisons with a standard, including the conditions under which discrepancies produce deleterious behavior, dysphoria, and self-concept change (see, e.g., Brickman & Janoff-Bulman, 1977; Campbell, 1986; Gibbons & Gerrard, 1991; Levine & Moreland, 1987; Miller & McFarland, 1991; Morse & Gergen, 1970; Wood, 1989; Wylie, 1979).

Investigation of the antecedents of comparative self-appraisal has proceeded in a much less systematic fashion. Researchers have examined various processes that are involved in the production of social norms and standards, but their investigations have been motivated by a disparate set of questions, many unrelated to the study of comparative appraisal. Nevertheless, we believe that the accumulated literature provides considerable insight into the processes that determine what serves as a standard of comparison. Our goal in this chapter is to substantiate that claim by examining two sets of processes: inference processes, through which people represent the attributes of individuals and of groups, and selection processes, through which they choose among potential sources of comparison. We focus our review on how these processes produce the standards used for comparative self-judgments. The psychological and behavioral consequences of the judgments are of interest here only insofar as they reflect and, in some cases, influence the choice of comparison target.

CONCEPTUALIZING SOCIAL NORMS AND STANDARDS

Although there is general agreement that social experience is evaluated comparatively, there is little agreement

Preparation of this chapter was supported by National Institute of Mental Health Grant MH44069. We thank Rick Gibbons, Len Newman, and Jacquie Vorauer for their helpful comments on an earlier draft of this chapter.

799

Psychological Methods and Simple Algorithms

- Historically: super controlled experiments on small samples of undergrads
  - Invite participants to the lab
  - Randomly assign them to one of two tasks
  - Collect data
  - Analyze data
  - Write an article

Standard Dataset

participant_id	reaction_time	task
1001	9395	1
1002	38902	1
1003	14538	0
1004	13918	0
1005	10925	1
1006	19169	1
1007	18052	1
1008	22744	0

Pick a method

- T-test, ANOVA, Simple linear regression
  - They all yield same results!
- Learn how to apply the method you picked

T-test in R

```
t.test(
  data$reaction_time[data$task == 0],
  data$reaction_time[data$task == 1],
  var.equal = TRUE)
```

ANOVA in R

```
TukeyHSD(
  aov(reaction_time ~ task,
      data = data)
  )
```

Simple Linear Regression in R

```r
summary(
  lm(reaction_time ~ task,
     data = data)
  )
```

That's it!!

You have acquired 90% of the knowledge you need to analyze experimental data from a super controlled lab experiment that went extremely well.

Problem?

Problem:

Resistance to Change and Need for Methodological Plasticity

Resistance to Change

- Simple Algorithms make it difficult to adopt:
    - New methods
    - New study designs
    - New technological tools

Methodological Plasticity

- Psychology researchers need **Methodological Plasticity**
- Methodological Plasticity can be defined as the ability to:
  - quickly grasp new statistical methods and research practices
  - make reasonable interpretations of a wide array of statistical analyses
  - feel comfortable adopting new analytic strategies and computational tools

Psychology Today

Many Methods

Changing climates of conflict: A social network experiment in 56 schools

Elizabeth Levy Paluck[a,1], Hana Shepherd[b], and Peter M. Aronow[c,d]

Evolution in Mind: Evolutionary Dynamics, Cognitive Processes, and Bayesian Inference

Jordan W. Suchow,[1,*] David D. Bourgin,[1] and Thomas L. Griffiths[1]

Many Methods

Computational Justice:
Simulating Structural Bias and Interventions

Ida Momennejad[1], Stacey Sinclair[2], Mina Cikara[3]
1 Columbia University, 2 Princeton University, 3 Harvard University

A Graph-Theoretic Approach to Multitasking

Jonathan D. Cohen * Biswadip Dey † Tom Griffiths ‡
Sebastian Musslick § Kayhan Özcimder ¶ Daniel Reichman ‖
Igor Shinkar ** Tal Wagner ††

Many Methods

Modeling the Partial Productivity of Constructions

Libby Barak and **Adele E. Goldberg**
Psychology Department
Princeton University
NJ, USA 08540

Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning

Tal Yarkoni and Jacob Westfall
University of Texas at Austin

Many Roles

- Evaluate research
  - Survey the literature
  - Feedback to colleagues
  - Attend talks
  - Collaborate
  - Review manuscripts

Many Roles

- Be evaluated by people who hold very different views!
  - Editors, reviewers, colleagues, students
  - Your methodological decisions will be challenged

I repeat, your methodological decisions will be challenged

Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis

Robin Gomila

Princeton University

Word count = 5,505

Logistic vs. Linear on Academic Twitter

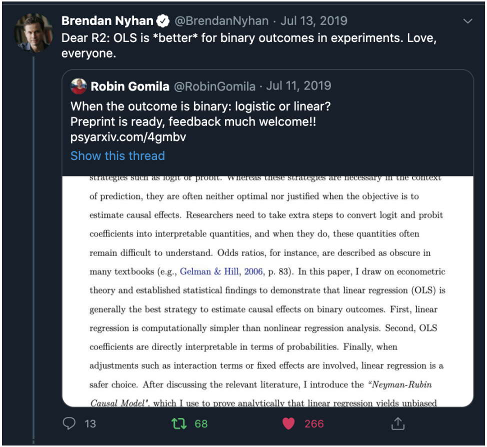
Robin Gomila
@RobinGomila

When the outcome is binary: logistic or linear?
Preprint is ready, feedback much welcome!!
psyarxiv.com/4gmbv

Abstract

When the outcome of interest is binary, psychologists often use nonlinear modeling strategies such as logit or probit. Whereas these strategies are necessary in the context of prediction, they are often neither optimal nor justified when the objective is to estimate causal effects. Researchers need to take extra steps to convert logit and probit coefficients into interpretable quantities, and when they do, these quantities often remain difficult to understand. Odds ratios, for instance, are described as obscure in many textbooks (e.g., Gelman & Hill, 2006, p. 83). In this paper, I draw on econometric theory and established statistical findings to demonstrate that linear regression (OLS) is generally the best strategy to estimate causal effects on binary outcomes. First, linear regression is computationally simpler than nonlinear regression analysis. Second, OLS coefficients are directly interpretable in terms of probabilities. Finally, when adjustments such as interaction terms or fixed effects are involved, linear regression is a safer choice. After discussing the relevant literature, I introduce the *"Neyman-Rubin Causal Model"*, which I use to prove analytically that linear regression yields unbiased estimates of causal effects, even when outcomes are binary. Then, I run simulations and analyze existing data on 24,191 students from 56 middle-schools (Paluck, Shepherd, & Aronow, 2016) to illustrate the effectiveness of linear regression with binary outcomes. Based on these grounds, I recommend that psychologists use linear regression instead of logit or probit models to estimate causal effects on binary outcomes.

Logistic vs. Linear on Academic Twitter



Logistic vs. Linear on Academic Twitter

Brendan Nyhan @BrendanNyhan · Jul 13, 2019

Dear R2: OLS is *better* for binary outcomes in experiments. Love, everyone.

> **Robin Gomila** @RobinGomila · Jul 11, 2019
>
> When the outcome is binary: logistic or linear?
> Preprint is ready, feedback much welcome!!
> psyarxiv.com/4gmbv
>
> Show this thread
>
> strategies such as logit or probit. Whereas these strategies are necessary in the context of prediction, they are often neither optimal nor justified when the objective is to estimate causal effects. Researchers need to take extra steps to convert logit and probit coefficients into interpretable quantities, and when they do, these quantities often remain difficult to understand. Odds ratios, for instance, are described as obscure in many textbooks (e.g., Gelman & Hill, 2006, p. 83). In this paper, I draw on econometric theory and established statistical findings to demonstrate that linear regression (OLS) is generally the best strategy to estimate causal effects on binary outcomes. First, linear regression is computationally simpler than nonlinear regression analysis. Second, OLS coefficients are directly interpretable in terms of probabilities. Finally, when adjustments such as interaction terms or fixed effects are involved, linear regression is a safer choice. After discussing the relevant literature, I introduce the *"Neyman-Rubin Causal Model"*, which I use to prove analytically that linear regression yields unbiased

13 · 68 · 266

Nate Silver @NateSilver538

Replying to @BrendanNyhan

Ehh, I kind of think this is wrong. If the underlying system your model describes is bounded between 0 and 1, your model probably should be also.

12:57 PM · Jul 13, 2019 · Twitter Web App

Debates and controversies around statistical methods happen all the time

SHARE

TECHNICAL COMMENTS

Comment on "Estimating the reproducibility of psychological science"

Daniel T. Gilbert[1,*,†], **Gary King**[1], **Stephen Pettigrew**[1], **Timothy D. Wilson**[2]

[1]*Harvard University, Cambridge, MA, USA.*

[2]*University of Virginia, Charlottesville, VA, USA.*

↵*Corresponding author. E-mail: gilbert@wjh.harvard.edu

– Hide authors and affiliations

Science 04 Mar 2016:
Vol. 351, Issue 6277, pp. 1037
DOI: 10.1126/science.aad7243

Article | Info & Metrics | eLetters | PDF

Abstract

A paper from the Open Science Collaboration (Research Articles, 28 August 2015, aac4716) attempting to replicate 100 published studies suggests that the reproducibility of psychological science is surprisingly low. We show that this article contains three statistical errors and provides no support for such a conclusion. Indeed, the data are consistent with the opposite conclusion, namely, that the reproducibility of psychological science is quite high.

Foundational Knowledge and Methodological Plasticity

- Empower you as a researcher
  - Many methods
  - Many roles
  - Many opinions

Building Plasticity

Yes, but how?

Underlying Structure of Psychological Methods

- Two components:
  - Statistical
  - Computational

Statistical Pillars of Psychological Methods

- Probability Theory
- Causality
- Regression

Probability Theory

Relevance

Evaluating a claim

- At a dinner party, your friend Charlie says:

  "I have a superpower. Since I was a kid, I've been able to see what people are doing without being with them. I just close my eyes, think about someone, and I find myself in the room with them. Let me try and find out what my sister is currently doing... At the moment, she is walking her dog."

- In search for evidence, you suggest:

  "Charlie, I am going to go to the kitchen to prepare coffee for everyone. I'll make eight cups of coffee, four of which will contain sugar. Will you be able to tell which cups contain sugar when I come back?"

- Charlie accepts the challenge and correctly identifies the cups that contain sugar

Evaluating a claim

- Should you conclude that Charlie has superpowers?

This is a serious question

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

0022-3514/11/$12.00 DOI: 10.1037/a0021524

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size (*d*) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about psi, issues of replication, and theories of psi are also discussed.

Scientific Approach: Does Charlie have superpowers?

- Scientific approach is probabilistic!
- Null Hypothesis
  - H_0: Charlie does not have special abilities
- Hypothesis testing
  - Does this hypothesis make sense, given the data? Do the data contradict this hypothesis, making it highly implausible?
  - Proof by contradiction

Testing the null hypothesis

- Under H_0: 70 possible outcomes
  - Charlie could have randomly chosen 4 cups in 70 different ways
  - Combinations of 4 possible elements within a broader set of 8 elements

$[1,2,3,4]$	$[1,2,5,7]$
$[1,2,3,5]$	$[1,2,5,8]$
$[1,2,3,6]$	$[1,2,6,8]$
$[1,2,3,7]$	$[1,2,7,8]$
$[1,2,3,8]$	$[1,3,4,5]$
$[1,2,4,5]$	$[1,3,4,6]$
$[1,2,4,6]$	$[1,3,4,7]$
$[1,2,4,7]$	$[1,3,4,8]$
$[1,2,4,8]$	$[1,3,5,6]$
$[1,2,5,6]$	$[1,3,5,7]$

$[\ldots]$

$[4,5,6,7]$	$[4,6,7,8]$
$[4,5,6,8]$	$[5,6,7,8]$

Testing the null hypothesis

- Under H_0, Charlie 's chances of finding the combination of 4 cups is:
  - $\frac{1}{70} = .014$
- This number is a probability:
  - Charlie has 1.4% chances of finding the correct combination without superpowers
- This number is a p-value
  - $p < .05$
  - Reject the null
  - Conclude that Charlie has superpowers

Lessons from Charlie

- Science is cumulative
  - Repeat performance several times
  - Fisher: *"Personally, the writer prefers to set a low standard of significance at the 5 percent point... A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance."*
- Psychological methods build heavily on *Probability Theory*
  - Probability theory is a framework for understanding and quantifying uncertainty
  - Psychology studies inevitably involve some degree of randomness and therefore, uncertainty

Causality

Relevance

What types of questions do psychological scientists ask?

- Psychologists (most often) ask causal questions
  - Are police officers more likely to shoot Black people?
  - Do diversity trainings reduce prejudice against minority groups?
  - Does social support improve well-being? Does cognitive-based therapy reduce depressive symptoms?
- Understand whether a variable X causes a psychological or behavioral outcome Y

Causality: a giant with feet of clay

- Human mind tends to interpret relationships in causal terms, and is often wrong! Examples?
  - People's race and their performance at "intelligence" tests
  - People's gender and mathematical abilities
  - Moving to California and life expectancy

Seriously, should we all move to California?

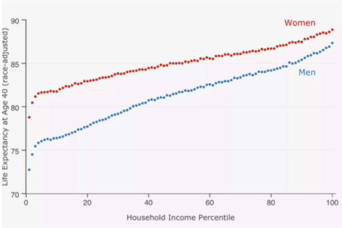
Vox

AD

Want to live longer, even if you're poor? Then move to a big city in California.

By Ezra Klein | @ezraklein | Apr 13, 2016, 1:30pm EDT

SHARE



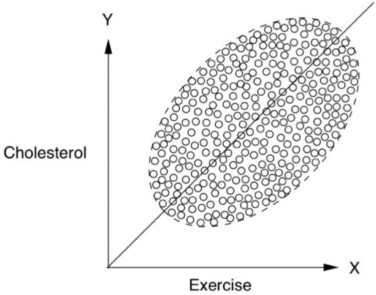
Journal of the American Medical Association

Causality: a giant with feet of clay

- Causality is difficult to establish
  - Rigorous studies and statistical analyses easily get it wrong.
  - Subtle elements of study designs, randomization procedures, participants' behavior during the study, analytic strategies, or statistical analyses can compromise researchers' ability to make a causal claim.

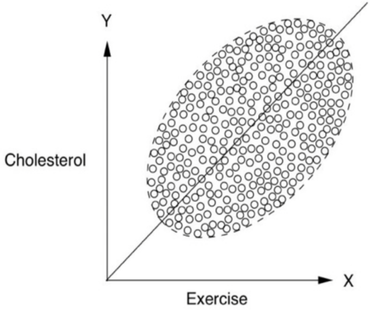
Causality: a giant with feet of clay

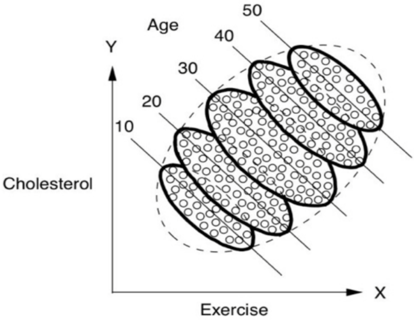
- Simpson's Paradox



Causality: a giant with feet of clay

- Simpson's Paradox





Regression

Relevance

What is regression?

- A powerful and highly-flexible tool to represent the relationship between variables
- Describe how an outcome Y varies with a single or a series of variables X
- Suited for the analysis of experimental, quasi-experimental and observational datasets, and can be used to make predictions.

Computational Foundations

Computational Tools

- Expect them to change