Statistics 209B / EPI 239 / Education 260A **Winter 2022**

Applications of Causal Inference Methods

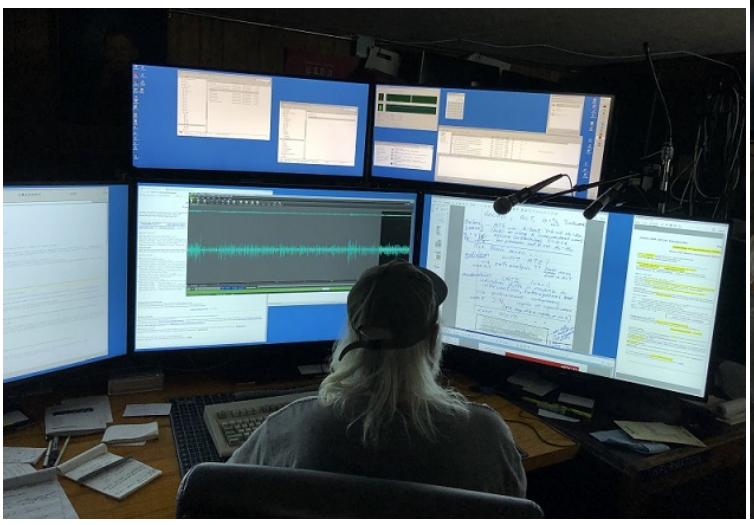
Winter 2022 Flipped Instruction

Instructor. David Rogosa, Sequoia 224, rag{AT}stanford{DOT}edu
TA David Fager dfager {AT} stanford {DOT} edu
Course web page: http://rogosateaching.com/stat209//

Course Welcome and Logistics (first day stuff, to be posted soon in December, call it Week0)
Lecture slides, week 0 (pdf) **Audio companion, week 0**

do this first

For recreation of in-classroom experience, linked below are youtube versions of the music I play
before starting lecture and after lecture concludes. Some may wish to reverse that ordering.



Registrar's Information

Statistics 209B (also EPI 239, EDUC 260A) 2 units
Title: Applications of Causal Inference Methods
Description: Application of potential outcomes formulation for causal inference to research settings including: mediation, compliance adjustments, time-1 time-2 designs, encouragement designs, heterogeneous treatment effects, aggregated data, instrumental variables, analysis of covariance regression adjustments, and implementations of matching methods.
see class website http://rogosateaching.com/stat209/
suggested prerequisite: an introduction to causal inference methods such as STAT209 (aut).

Course Overview
Note: This course was structured before the new world order of March 2020 as one two-hour lecture class meeting per week.
Last year, Winter 2021, this course was done Remote Asynchronous. Lectures (pdf slides plus audio) and all supporting materials are online.
To see full course materials from Winter 2021 go here
This year, the course will be done in flipped mode, with lectures (pdf slides and audio) and all supporting materials linked in these pages.
Web page materials to be posted sequentially for that week's instruction-- i.e. Week 1 materials to be posted Jan 3, and so forth.
The Wednesday class meeting, not required, will be used in part for review and discussion of the prior weeks' materials.
More on this in the Week0 intro materials.
The **~~first~~ in-class discussion section ~~will be January 12, 2022~~ to be determined by health conditions**.

Brief Course Outline

Unit 1. Extensions of RCT
Analyzing Encouragement Designs
Assessing Mediation in experimental studies
Identifying Moderation in experimental studies (heterogeneous treatment effects)
The wisdom of Compliance Adjustments (for binary and measured compliance);
Analysis of Regression Discontinuity Designs (systematic assignment based on a covariate)
Unit 2. Legacy Methods for Causal Inference from Observational Data (good and bad)
Regression adjustments (analysis of covariance) in observational studies
Instrumental Variables methods for observational studies
Failures of traditional Path Analysis (and Structural Equation Models)
Interpreting associations: Spurious Correlation and Simpson's Paradox
Multilevel data and Contextual Effects
Reciprocal Causal Effects and non-recursive models
Unit 3. Historical and Modern Methods for Matching in Observational Data
Case-Control, case-referent matching sudies
Subclassification matching
Propensity Score matching examples
Unit 4. Time-1, Time-2 data in experimental and observational studies.

Experimental Designs	Cross-over designs
Experimental Designs	Comparing groups on time-1, time-2 measurements: repeated measures anova vs lmer OR the t-test
Observational studies	Economist's differences in differences (or diffs in diffs with matching) for observational studies.
Observational studies	Lord's paradox; pre-post group comparisons.
Observational studies	Exogenous Variables and Correlates of Change (use of lagged dependent variables)
Additional Special topics	Interrupted Time-series designs; Current implementations of value-added analysis

Lectures, Course Files, and Readings
this page is where course content resides

Grading, Exams, and Credit Units
Stat209B/EPI239/Ed260A is listed as Letter or Credit/No Credit grading for 2-units
For Winter 2022 grading for the 2-units will be based on a 'take home'(i.e. do at home) Problem Set.
Also as you will see (or can view from 2021 pages), for each week's content a number of Review Questions with Solutions are posted.

Course Problem Set 2022 to be posted

Statistical computing
Class presentation will be in, and students are encouraged to use, R (occasionally, some references to SAS and Mathematica).
Current version of R is R version 4.1.2 released 2021-11-01
For references and software: The R Project for Statistical Computing Closest download mirrors in the past, UCLA and Berkeley, seem no longer avaliable, pick your fave anywhere in the world.
The CRAN Task Views provide useful overviews/navigation for the almost infinite set of R-packages. Task Views that cover contents of this course include CRAN Task View: Statistics for the Social Sciences; CRAN Task View: Econometrics and CRAN Task View: Psychometric Models and Methods .

Relevent Texts (optional).
Causal Inference in Statistics, Social and Biomedical Sciences: An Introduction, Guido Imbens and Don Rubin, 1st Edition (Cambridge University Press) Stanford access
Design of Observational Studies, Paul Rosenbaum, 1st Edition (Springer). Available online: Stanford access

---

To see full course materials from legacy Stat209 (2005-2019)
Statistical Methods for Group Comparisons and Causal Inference go here

---

Stat 209B-- Lectures, Course Files, and Readings

Week 0

Course introduction (slides and audio ~~to be~~ posted on main page)

Background readings (not required, but of interest if you haven't seen these before)

1. Correlation and Causation: A Comment, Stephen Stigler *Perspectives in Biology and Medicine*, volume 48, number 1 supplement (winter 2005)
2. Secret to Winning a Nobel Prize? Eat More Chocolate (Time)
   Publication: Chocolate Consumption, Cognitive Function, and Nobel Laureates Franz H. Messerli, M.D. N Engl J Med 2012; 367:1562-1564 October 18, 2012
3. *David Freedman chapters*.
   From Association to Causation: Some Remarks on the History of Statistics;
   Statistical Models for Causation: A critical review
   Statistical Models and Shoe Leather, *Sociological Methodology*, Vol. 21. (1991), pp. 291-313. JStor link

Week 1

~~To be posted~~
Lecture slides, week 1 (pdf)
Audio companion, week 1
parta partb partc

1. Encouragement Designs: example of potential outcomes formulation.

Lecture Topics

Illustration using encouragement design representation in Holland (1988). copies of selected overheads.
Encouragement Designs. Potential outcomes formulation and IV parameter estimation in Holland (1988). Estimation handout
Do regression methods (path analysis) identify causal effects? Demonstrations of failure for Holland's encouragement design. class handout Encouragement design slides

Primary Readings

Paul Holland, Causal Effects and Encouragement Designs. Causal Inference, Path Analysis, and Recursive Structural Equations Models
Paul W. Holland Sociological Methodology, Vol. 18. (1988), pp. 449-484. (Encouragement design results; sections 3-5)
Holland Appendix (esp pp. 475-480) presents the potential outcomes formulation.

Abstract Rubin's model for causal inference in experiments and observational studies is enlarged to analyze the problem of "causes causing causes" and is compared to path analysis and recursive structural equations models. A special quasi-experimental design, the encouragement design, is used to give concreteness to the discussion by focusing on the simplest problem that involves both direct and indirect causation. It is shown that Rubin's model extends easily to this situation and specifies conditions under which the parameters of path analysis and recursive structural equations models have causal interpretations.

Encouragement Design research examples:

Sesamee Street evaluation
Gelman-Hill text sec 10.5; Data Analysis Using Regression and Multilevel/Hierarchical Models
Salt and Blood Pressure clinical trial
Publication: Feasibility and efficacy of sodium reduction in the Trials of Hypertension Prevention, phase I Trials of Hypertension Prevention Collaborative Research Group. S K Kumanyika, P R Hebert, J A Cutler, V I Lasser, C P Sugars, L Steffen-Batey, A A Brewer, MI. Hypertension doi: 10.1161/01.HYP.22.4.5021993;22:502-512

2. Mediating (process) variables

Lecture Topics

Historical (Barron-Kenny) methods David Kenny web page
R-implementations: mediating variables data analysis example data file
Barron-Kenny method via Sobel function in the multilevel package.
More extensive implementation (incl BCa bootstrapping) function `mediation` in package `MBESS` Ken Kelley;
power and sample size calculations in package `powerMediation`
`mediation` package. takes the topic up a large level of complexity/capabilities

Primary Readings

Vignette for `mediation` package Causal Mediation Analysis Using R
Mediation Analysis David P. MacKinnon, Amanda J. Fairchild, and Matthew S. Fritz Department of Psychology, Arizona State University, Tempe, Arizona 85287-1104; Annu. Rev. Psychol. 2007. 58:593-614

Mediation research examples:

Framing experiment
Brader T, Valentino NA, Suhat E (2008). What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration." American Journal of Political Science, 52(4), 959-978. jstor link
Data in `mediation` package; data description and analyses in `mediation` package vignette (linked below)
Bench Science vs Path Analysis: Exercise and Alzheimers
The irisin bench-science mediation example is discussed at the beginning of Week 2 lecture for recap and because I couldn't find it at the time.
NYTimes:How Exercise May Help Keep Our Memory Sharp .
Publication: Exercise-linked FNDC5/irisin rescues synaptic plasticity and memory defects in Alzheimer's models *Nature Medicine* volume 25, pages165-175 (2019)
Mediated moderation?
Stanford Medicine Common opioids less effective for patients on SSRI antidepressants Publication: Predicting inadequate postoperative pain management in depressed patients: A machine learning approach Arjun Parthipan,Imon Banerjee,Keith Humphreys,Steven M. Asch,Catherine Curtin,Ian Carroll ,Tina Hernandez-Boussard Published: February 6, 2019https://doi.org/10.1371/journal.pone.0210575
New Yorker. December 23, 2013. The Power of the Hoodie-Wearing C.E.O. Publication: The Red Sneakers Effect: Inferring Status and Competence from Signals of Nonconformity Author(s): Silvia Bellezza, Francesca Gino, and Anat Keinan Source: *Journal of Consumer Research*

Additional Resources

Mediators and Moderators of Treatment Effects in Randomized Clinical Trials Helena Chmura Kraemer; G. Terence Wilson; Christopher G. Fairburn; W. Stewart Agras Arch Gen Psychiatry. 2002;59:877-883
additional technical papers. Causal Mediation Analysis Using R K. Imai, L. Keele, D. Tingley, and T. Yamamoto American Political Science Review Vol. 105, No. 4 November 2011
Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies
MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M.,West, S. G., Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7, 83-104.
Useful expositions Using R
Chapter 14: Mediation and Moderation Alyssa Blair
Mediation and Moderation Analyses with R - OSF presentation slides

Week 1 Review Questions

Question 1. Mediating Variable Computations: Class example continued

The data set shown in class example ss423 is linked above and in the legacy directory http://web.stanford.edu/~rag/stat209/ss423
for predictor (IV) 'belong' outcome 'depress' and (potential) mediating variable 'master' The class example showed you the Baron-Kenny analysis using functions from the multilevel and MBESS packages.
Here just use 'lm' basic regression and the recipees from the class handout to recreate point estimates and asymptotic standard errors, significance tests for the mediating variable effect.
Compare your result with the class example posting.
Extra: also try out the more 'sophisticated' functions in the mediation package.

Solution for question 1

Question 2. Potential Outcomes, Encouragement Design Estimation and (Causal) Mediation

Task 1. Create a potential Outcomes dataset following the first ALICE specification in the posted slides (week 3) ## ALICE example beta = 3 rho = 3 tau = 1, delta = 3 (I did n=400; larger would be better so I redid with n = 6400)

Task 2. Use the artificial data to show the results for the mediation (indirect) effect by hand doing the 3 regressions using multilevel package (sobel) using MBESS package using the causal mediation estimation ACME from the mediation package and compare with rho*beta

Task 3 estimate beta by the Wald estimator (assuming tau = 0) and estimate mediation effect

Solution for question 2

Question 3. Sesame Street: Encouragement Design research example

Sesame Street research setting and data description given pdf p.30 of Lecture 1 (also Gelman text).
For this exercise use `postnumb` : posttest on numbers (0-54), along with the measures encour and regular from the class example in Lecture 1.
Use the encouragement design formulation to estimate the effect on child cognitive development (postnumb here) of watching more Sesame Street.
What assumption is necessary for the IV estimation in this design?
Obtain a point and interval estimate for the effect of viewing (use `ivreg` as in class example).
From simple descriptives reproduce this instrumental variables estimate (Wald estimator).
The second approach (path analysis) analyzed by Holland requires what assumption?
Obtain the path analyses (regression) estimate for the effect on child cognitive development (postnumb here) of watching more Sesame Street.
Compare with the IV estimate (which employs different assumptions).

Solution for question 3

Week 2

Moderating Variables in experimental studies (heterogeneous treatment effects)

13

Causal Inference, Path Analysis, and Recursive Structural Equations Models

*Paul W. Holland**

Rubin's model for causal inference in experiments and observational studies is enlarged to analyze the problem of "causes causing causes" and is compared to path analysis and recursive structural equations models. A special quasi-experimental design, the encouragement design, is used to give concreteness to the discussion by focusing on the simplest problem that involves both direct and indirect causation. It is shown that Rubin's model extends easily to this situation and specifies conditions under which the parameters of path analysis and recursive structural equations models have causal interpretations.

1. INTRODUCTION

The perspective on causal inference developed extensively by Rubin (1974, 1977, 1978, 1980, 1986) provides a solid basis for considering issues of causal inference in complex cases, and it is the only one that is grounded where causal inferences are relatively uncontroversial—in experimental science. In Holland (1986*a*, *b*), I described this perspective and dubbed it *Rubin's model*, as I will refer to it here, too, even though a more general term, such as *the experimental model*,

I thank David Freedman, Clark Glymour, Edward Leamer, Margaret Marini, James Robins, David Rogosa, Donald Rubin, Burton Singer, and Howard Wainer for their many helpful comments on an earlier draft of this paper.

*Educational Testing Service.

Encouragement Designs

Holland (1988) Powers & Swinton (1984)

Intervention: Encouragement to study (for a test)

random assignment to treatment - control
G=1 G=0

Student studies amount R

Student outcome, achievement test score, Y

For each unit observe:

Y, R, G

Other examples: Behavioral Medicine, salt ex

Hypertension
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Feasibility and efficacy of sodium reduction in the Trials of Hypertension Prevention, phase I. Trials of Hypertension Prevention Collaborative Research Group

SK Kumanyika, PR Hebert, JA Cutler, VI Lasser, CP Sugars, L Steffen-Batey, AA Brewer, M Cameron, LD Shepek and NR Cook

Hypertension 1993;22;502-512
Hypertension is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514

 Print ISSN: 0194-911X. Online ISSN: 1524-4563

Feasibility and Efficacy of Sodium Reduction in the Trials of Hypertension Prevention, Phase I

Shiriki K. Kumanyika, Patricia R. Hebert, Jeffrey A. Cutler, Vera I. Lasser, Carolyn P. Sugars, Lyn Steffen-Batey, Amy A. Brewer, Mary Cameron, Lana D. Shepek, Nancy R. Cook, Stephen T. Miller
for the Trials of Hypertension Prevention Collaborative Research Group

Phase I of the Trials of Hypertension Prevention was a multicenter, randomized trial of the feasibility and efficacy of seven nonpharmacologic interventions, including sodium reduction, in lowering blood pressure in 30- to 54-year-old individuals with a diastolic blood pressure of 80 to 89 mm Hg. Six centers tested an intervention designed to reduce dietary sodium to 80 mmol (1800 mg)/24 h with a total of 327 active intervention and 417 control subjects. The intervention consisted of eight group and two one-to-one meetings during the first 3 months, followed by less-intensive counseling and support for the duration of the study. The mean net decrease in sodium excretion was 43.9 mmol/24 h at 18 months. Women had lower sodium intake at baseline and were therefore more likely to decrease to less than 80 mmol/24 h. Black subjects were less likely to decrease to less than 80 mmol/d, independent of sex or baseline sodium excretion. The mean (95% confidence interval) net decrease associated with treatment was −2.1 (−3.3, −0.8) mm Hg for systolic blood pressure and −1.2 (−2.0, −0.3) mm Hg for diastolic blood pressure at 18 months (both $P<.01$). Multivariate analyses indicated a larger systolic blood pressure effect in women (−4.44 versus −1.23 mm Hg in men), adjusted for age, race, baseline blood pressure, and baseline 24-hour urinary sodium excretion ($P=.02$). Dose-response analyses indicated an adjusted decrease of −1.4 mm Hg for systolic blood pressure and −0.9 mm Hg for diastolic blood pressure for a decrease of 100 mmol/24 h in 18-month sodium excretion. These results support the utility of sodium reduction as a population strategy for hypertension prevention and raise questions about possible differences in dose response associated with gender and initial level of sodium intake. (*Hypertension*. 1993;22:502-512.)

KEY WORDS • **hypertension, sodium-dependent • blood pressure • sodium, dietary • primary prevention • blacks • women**

Primary prevention of hypertension, ie, preventing people at risk of hypertension from developing it, could potentially lower death rates from cardiovascular disease, reduce the need for antihypertensive medications, and reduce hypertension-related medical costs and job absenteeism.[1-4] However, the feasibility and efficacy of sodium reduction in lowering blood pressure among people with normal blood pressure in the general population have not been clearly established. Several lines of evidence suggest that sodium reduction is a logical candidate for incorporation into a primary prevention strategy. The well-established blood pressure–lowering effect of sodium reduction as a component of hypertension treatment is not confined to blood pressures above the physiologically arbitrary cutoffs used to define high blood pressure. Small, controlled trials have demonstrated the efficacy of moderate sodium reduction in reducing blood pressure in normotensive and hypertensive people.[5,6] In addition, data from the INTERSALT study have confirmed that, over a wide range of sodium intake, populations with low sodium consumption have lower blood pressures than those with high sodium consumption.[7,8]

Phase I of the Trials of Hypertension Prevention (TOHP-I) was a multicenter, randomized trial designed to test the short-term efficacy and safety of several nonpharmacologic interventions, including sodium reduction, in reducing blood pressure in a large cohort of men and women without hypertension in the context of their usual patterns of living.[9] TOHP-I attempted to provide an unambiguous answer regarding the potential for lowering the average blood pressure levels in the

Received December 28, 1992; accepted in revised form June 4, 1993.

From the Department of Epidemiology, The Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md (S.K.K., L.S.-B.); Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Mass (P.R.H., N.R.C.); Prevention and Demonstration Research Branch, Division of Epidemiology and Clinical Applications, National Heart, Lung, and Blood Institute, Bethesda, Md (J.A.C.); Preventive Cardiology Program, Department of Medicine, New Jersey Medical School, Newark (V.I.L.); Department of Internal Medicine, University of California, Davis (C.P.S.); Department of Preventive Medicine, University of Tennessee, Memphis (A.A.B., S.T.M.); University of Mississippi, Jackson, (M.C.); and St Louis University School of Medicine (Mo) (L.D.S.).

A preliminary version of these results was presented at the 64th Scientific Sessions of the American Heart Association, Anaheim, Calif, November 1991, and the International Heart Health Conference, Vancouver, British Columbia, Canada, May 1992.

Correspondence to Shiriki Kumanyika, Center for Biostatistics and Epidemiology, College of Medicine, Pennsylvania State University, PO Box 850, Hershey, PA 17033.

The Sesame Street example is from Andrew Gelman at Columbia,
We consider data from an experiment in which a randomly selected group of children was encouraged to watch the television program Sesame Street and the randomly selected control group was not. The goal of the experiment was to estimate the effect on child cognitive development of watching more Sesame Street. In the experiment, encouragement, but not actual watching, was randomized.

We might consider implementing a randomized experiment where the participants are preschool children, the treatment of interest is watching Sesame Street, the control condition is not watching, and the outcome is the score on a test of letter recognition. It is not possible here for the experimenter to force children to watch a TV show or to refrain from watching (the experiment took place while Sesame Street was on the air). Thus watching cannot be randomized. Instead, when this study was actually performed, what was randomized was encouragement to watch the show--this is called a randomized encouragement design

```
-------------------------
Code book with variable names for Sesame Street data
id : subject identification number
site :   1 =Three to five year old disadvantaged children from inner city areas in
              various parts of the country.
           2 = Four year old advantaged suburban children.
           3 = Advantaged rural children.
           4 = Disadvantaged rural children.
           5 = Disadvantaged Spanish speaking children.
sex   male=1, female=2
age   age in months
viewcat  frequency of viewing
              1=rarely watched the show
              2=once or twice a week
              3=three to five times a week
              4=watched the show on average more than 5 times a week
setting:    setting in which Sesame Street was viewed, 1=home 2=school
viewenc :  treatment condition
            1=child encouraged to watch,  2=child not encouraged to watch
encour:  treatment condition
           0=child not encouraged to watch, 1=child encouraged to watch
regular:  frequency of viewing:
          0=rarely watched the show, 1= watched once/week or greater
prebody :  pretest on knowledge of body parts (scores range from 0-32)
prelet :  pretest on letters (scores range from 0-58)
preform : pretest on forms (scores range from 0-20)
prenumb : pretest on numbers (scores range from 0-54)
prerelat : pretest on relational terms (scores range from 0-17)
preclasf : pretest on classification skills
postbody : posttest on knowledge of body parts (0-32)
postlet :  posttest on letters (0-58)
postform :  posttest on forms (0-20)
postnumb : posttest on numbers (0-54)
postrelat : posttest on relational terms (0-17)
postclasf:  posttest on classification skills
peabody:  mental age score obtained from administration of the Peabody Picture
           Vocabulary test as a pretest measure of vocabulary maturity
{note: measures used in the analyses below are:
   encour, regular, postlet, prelet, site, peabody
  encour and viewenc are redundent recodings)
```

I have a data frame called "ses" ('cuz I can't spell sesame consistently)

```
> dim(ses) #we have 240 children
[1] 240  28
```

The data frame is in wide form (each child is a row)

```
> names(ses) #approximately correspond to the codebook listing
 [1] "rownames"  "id"        "site"      "sex"       "age"       "viewcat"   "setting"
     "viewenc"   "prebody"   "prelet"
[11] "preform"   "prenumb"   "prerelat"  "preclasf"  "postbody"  "postlet"   "postform"
     "postnumb"  "postrelat" "postclasf"
[21] "peabody"   "agecat"    "encour"    "_Isite_2"  "_Isite_3"  "_Isite_4"  "_Isite_5"
     "regular"
```

Other Examples of Encouragement Experiments:

Dylan Small

Outcome	Treatment	Instrumental Variable	Reference
Earnings	Participation in federal job training program	Random assignment of admission to training program	Bloom et al. (1997)
Achievement test scores	Enrollment in private school	Randomly selected offer of school voucher	Howell et al. (2000)
Achievement test scores	Class size	Random assignment to a small or normal-size class	Krueger (1999)
Depression Level	Meeting with depression specialist	Random encouragement to meet with depression specialist	Small et al. (2007)
Achievement test scores	Hours of study	Random mailing of test preparation materials	Powers and Swinton (1984)

Causal Effect Being Estimated When Treatment Effect is Not Constant

Suppose that we have an IV that satisfies the exclusion restriction but treatment effects are not constan

$$r_{(z,s)i} = r_{(0,0)i} + \beta_i s$$

Consider the case of a binary treatment S. We can divide the subjects into four groups:

(1) $s_{(1)i} = 0,\ s_{(0)i} = 0$ (Never takers): subjects who would take not take treatment, regardless whether encouraged to do so

- Questions

1. What's the increase in study time from Encouragement?

2. What's the increase in achievement from studying an hour longer?

3. What's the increase in achievement if no increase in study? (placebo effect: concentration, motivation, etc.)

4. What's the total impact on achievement?

Counterfactual Data for Individual u

with parameterization

1. Difference in amount of study time for u when in treatment vs when in control:

$$R_t(u) - R_c(u) = \rho(u)$$

2. Difference in outcome score for studying amount r vs studying amount r':

$$Y_{Gr}(u) - Y_{Gr'}(u) = (r - r')\beta(u)$$

3. Difference in outcome score when in treatment vs when in control with same amount of study r:

$$Y_{tr}(u) - Y_{cr}(u) = \tau(u)$$

4. Overall difference in outcome score when in treatment vs when in control:

$$Y_{tR_t}(u) - Y_{cR_c}(u)$$
$$= \tau(u) + \rho(u)\beta(u) \quad \text{(Direct + indirect)}$$

Formulation For Encouragement Designs

"Counterfactual Data"

1. $R_t(u) - R_c(u) = \rho(u)$

2. $Y_{Gr}(u) - Y_{Gr'}(u) = (r - r')\beta(u)$

3. $Y_{tr}(u) - Y_{cr}(u) = \gamma(u)$

4. $Y_{tR_t}(u) - Y_{cR_c}(u) = \gamma(u) + \rho(u)\beta(u)$

DIRECT + INDIRECT

4 ACE's: Expectation over units

Formulati[redacted] Encouragemen[t] Designs

"Counterfactual Data"

ALICE

1. $R_t(u) - R_c(u)$ — ρ
2. $Y_{Gr}(u) - Y_{Gr'}(u)$ — $\beta(r-r')$
3. $Y_{tr}(u) - Y_{cr}(u)$ — τ
4. $Y_{tR_t}(u) - Y_{cR_c}(u)$ — $\tau + \rho\beta$

4 ACE's over

Unit (Individual) Level Model

Amount of study:

$$R_G(u) = R_c(u) + \rho(u)G$$

Outcome score:

$$Y_{Gr}(u) = Y_{co}(u) + \gamma(u)G + \beta(u)r$$

↑

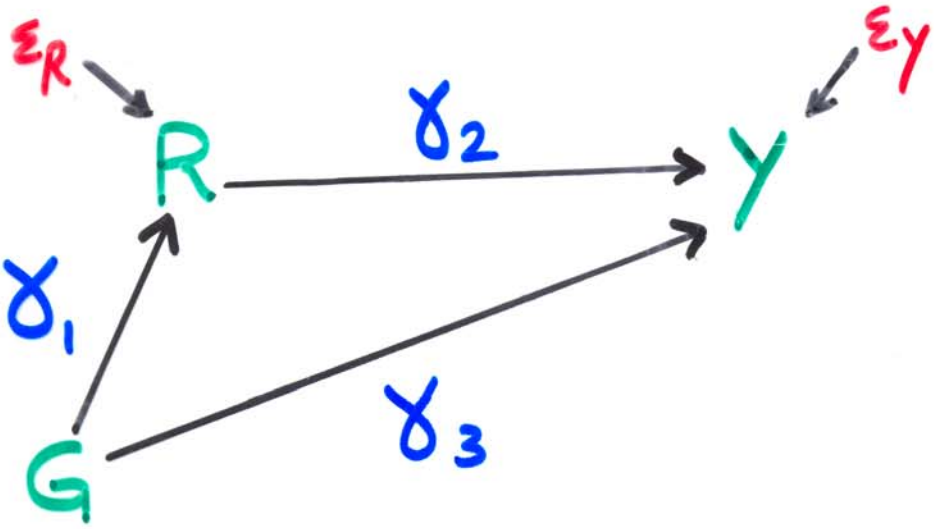
Outcome score of u if not encouraged and studies 0 time.

Simplest model (Holland, 1988)

No individual differences

$$\rho(u) = \rho \; ; \; \beta(u) = \beta \; ; \; \gamma(u) = \gamma$$

Path Analysis for Y, R, G



Path Regressions

R on G

Y on R, G

Path Analysis Results under ALICE

$\gamma_1 = \rho$ $\quad G \rightarrow R$

$\gamma_2 = \beta + \delta$ $\quad R \rightarrow Y$

$\gamma_3 = \tau - \rho\delta$ $\quad G \rightarrow Y$

For $\delta = \text{cov}(Y_{c0}(u), R_c(u)) / \text{VAR}(R_c(u))$

Causal Effects	ALICE	Path Analysis
Overall	$\tau + \beta\rho$	$\gamma_3 + \gamma_1\gamma_2 =$ $\tau + \beta\rho$ ✓
Indirect / Direct	$\beta\rho / \tau$	$\gamma_1\gamma_2 / \gamma_3 =$ $\rho(\beta+\delta)/(\tau-\rho\delta)$

[redacted]

Gold Standard Tests for Structural Equation Models using individual unit models (response functions)

A. Holland (1988) Encouragement Designs

individual unit model: (potential outcomes)

$R_G(u) = R_c(u) + \rho(u)\, G$; $Y_{Gr}(u) = Y_{c0}(u) + \tau(u) G + \beta(u) r$

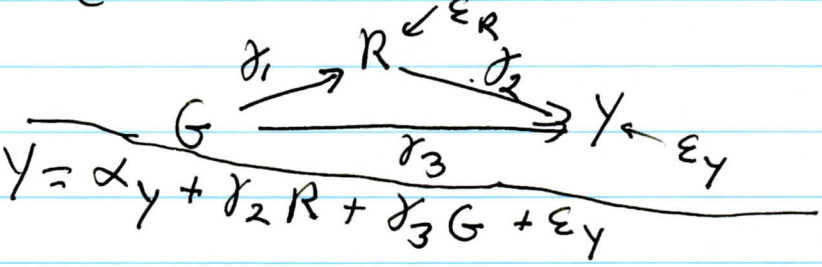
params ρ β τ

Path Analysis

using obs on G R Y

$R = \alpha_R + \gamma_1 G + \varepsilon_R$

$Y = \alpha_Y + \gamma_2 R + \gamma_3 G + \varepsilon_Y$



Compare γ_1 γ_2 γ_3

with $\rho(u)$ $\beta(u)$ $\tau(u)$

Under ALICE $\rho(u) = \rho$ $\beta(u) = \beta$ $\tau(u) = \tau$

results $\gamma_1 = \rho$ $\gamma_2 = \beta + \delta$ $\gamma_3 = \tau - \rho\delta$

bias due to $\delta = \mathrm{cov}(Y_{c0}(u), R_c(u)) / \mathrm{Var}(R_c(u))$

individual diff's matter

overall effect on Y $\tau + \beta\rho = \gamma_3 + \gamma_1\gamma_2$

direct/indirect $= \beta\rho/\tau$ vs $\frac{\rho(\beta+\delta)}{\tau - \rho\delta}$ path analysis

Numerical Examples (slides)

set $\beta = 3$ $\tau = 1$ $\mathrm{Corr}(R_c, Y_{c0}) = .75$, $\mathrm{Var}(Y_{c0}) = 64$

for $\rho = 3$, $\mathrm{Var}\, R_c = 4$)

$\gamma_1 = 3$ $\gamma_2 = 6$ $\gamma_3 = -8$

ind/direct $= -2.25$ (vs 9)

more exs on slides

Does it work?

Stat209

Path Analysis and Encouragement Design

drr summary

Encouragement Designs: Effects of Interventions

Exemplar Study: Random assignment of students to treatment-control conditions for intervention on improving study habits. *Measures:* Treatment/control assignment (G), amount of study (R), and outcome measure, achievement test score (Y).

Questions: **1.** Increase in study time from intervention? **2.** Increase in achievement from studying an hour longer (dose response)? **3.** Increase in achievement if no increase in study (placebo effect)? **4.** Total impact on achievement?

Path Analysis Regressions
Path Coefficients: $\gamma_1, \gamma_2, \gamma_3$

$R = \alpha_R + \gamma_1 G + \varepsilon_R$

$Y = \alpha_Y + \gamma_2 R + \gamma_3 G + \varepsilon_Y$

Counterfactual Data Formulation for Individual u. **1.** $R_t(u) - R_c(u) = \rho(u)$, treatment/control difference in amount of study. **2.** $Y_{Gr}(u) - Y_{Gr'}(u) = \beta(u)*(r - r')$, increment to outcome from study amount r' vs r. **3.** $Y_{tr}(u) - Y_{cr}(u) = \tau(u)$, treatment/control difference in outcome with same amount of study r. **4.** $Y_{tR}(u) - Y_{cR}(u) = \tau(u) + \rho(u)\beta(u)$, overall treatment/control difference.

Individual Level Model. $R_G(u) = R_c(u) + \rho(u)G$; $Y_{Gr}(u) = Y_{c0}(u) + \tau(u)G + \beta(u)r$.

ALICE specification: $\rho(u) = \rho$; $\beta(u) = \beta$; $\tau(u) = \tau$.

Path Analysis Results Under ALICE. $\gamma_1 = \rho$; $\gamma_2 = \beta + \delta$; $\gamma_3 = \tau - \rho\delta$.

Indirect/Direct Effects: $\beta\rho/\tau$ under ALICE; $\rho(\beta + \delta)/(\tau - \rho\delta)$ from path analysis.

Holland (1988) p. 469

$$\mu_c(r) = \gamma + \delta r. \tag{51}$$

A positive δ means that the more a student would study when not encouraged, the *higher* he or she would score on the test without studying and without encouragement. A negative δ means that the more a student would study when not encouraged, the *lower* he or she would score without studying and without encouragement.

S = 0/1 (G)

The quantities computed in path analysis are the conditional expectations

$$E(R_S|S) = E(R_c) + \rho S \tag{52}$$

and

$$E(Y_{SR_S}|S, R_S) = \mu_c(R_S - \rho S) + \tau S + \beta R_S, \tag{53}$$

in which S is a 1/0 indicator variable. If we make the untestable assumption that $\mu_c(r)$ is linear, e.g. (51), then (53) becomes

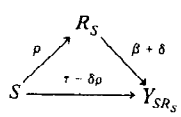
$$E(Y_{SR_S}|S, R_S) = \gamma + (\tau - \delta\rho)S + (\beta + \delta)R_S. \tag{54}$$

Equations (52) and (54) are both linear and may be combined into the empirical path diagram in Figure 5.

Comparing Figures 5 and 4, we see that even if the ALICE model holds and $\mu_c(r)$ is linear, the estimated path coefficients are

meaning of δ individual differences in outcome if null R, G

FIGURE 5.



biased estimates of the causal effects τ and β unless $\mu_c(r)$ does not depend on r (i.e. $\delta = 0$). Furthermore, these problems stem from the inhomogeneity of the units with respect to the values of $R_c(u)$ and $Y_{c0}(u)$. This inhomogeneity is, I believe, the proper way to view the "disturbance terms" of the structural equations model (9) and (10) in section 3 (see section 4.4). One nice thing is that while the *direct effects* are not the same in Figures 5 and 4, the total effects are: Both equal $\tau + \rho\beta$.

4.3. *Two Different Ways to Estimate the Causal Effect of the Encouraged Activity*

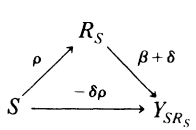
The message of the previous subsection is that the effect of study on test performance cannot be estimated by the usual regression methods of path analysis without making untestable assumptions about the counterfactual regression function, $\mu_c(r)$. If we assume that $\mu_c(r)$ is constant, then the biases shown in Figure 5 vanish and the usual path coefficients may be interpreted as causal effects, i.e. as ACEs. However, because $\mu_c(r)$ is so difficult to think about, there is little reason to believe that it is constant. Nor can δ be easily assessed as either positive or negative, since, in this example, there are reasons why it might be either: If students who study a lot tend to be those who do well even when they don't study, then δ is positive; but if those who study a lot are those who need to study, then δ is negative.

An alternative approach is to suppose that encouragement, of and by itself, has no effect on Y. In the studying example, this might be a plausible assumption. This corresponds to the restriction that

$$\tau = 0. \tag{55}$$

Now the empirical path diagram becomes that in Figure 6, and the path diagram for the causal theory becomes that in Figure 7. The total effect of S on Y_{SR_S} is now $\rho\beta$, whereas the total effect of S on R_S is ρ.

FIGURE 6.



Examples under ALICE

$\beta = 3$, $\tau = 1$, $Corr(R_c, Y_{co}) = .75$

$Var(Y_{co} = 64)$ $Var(R_c = 4)$

	$\rho = 1$ (.5 sd effect)	$\rho = 3$ (1+ sd effect)
G→R γ_1	1	3
R→Y γ_2	6	6
G→Y γ_3	-2	-8
ALICE ind/dir	3	9
PATH ind/dir	-3	-2.25

Even with random assignment to and control, path regressions do not recover the individual level model.

Example:

For individual level model:

$\beta(u) \sim U[2,8]$ $\gamma = 2$ $\rho = 8$

$\mu_\beta = 5$

Path analysis regressions give ($n = 10{,}000$)

$\hat{\gamma}_2 = 8.57$ $\hat{\gamma}_3 = -26.3$ $\hat{\gamma}_1 = 8.01$

Ratio of "indirect" to "direct" effects:

$\frac{\mu_\beta \rho}{\gamma} = 20$ $\frac{\hat{\gamma}_1 \hat{\gamma}_2}{\hat{\gamma}_3} = -2.6$

Artificial Data (Quadratic Response)

	Y	R	G	Variance
Y	1			2934.
R	.86	1		17.8
G	.78	.94	1	

Path Analysis gives:

$\hat{\gamma}_2 = 14.37$

$\hat{\gamma}_3 = -29.4$

$\hat{\gamma}_1 = 7.95$

Data generation parameters

$\mu_{\beta_1} + \mu_{\beta_2}\bar{R} = 6.8$

$\beta_1 \sim U[2,4]$; $\beta_2 \sim U[.1,.3]$

$\gamma = 2$

$\rho = 8$

Quadratic Response

$$Y(u) = Y_0(u) + \beta_1(u)R(u) + \beta_2(u)[R(u)]^2$$

More typical application in nonexperimental setting.

Level of encouragement (G) determined by home environment, i.e. outside experimental control.

Failure of path analysis in a simple experiment should strike fear into such applications.

biased estimates of the causal effects τ and β unless $\mu_c(r)$ does not depend on r (i.e. $\delta = 0$). Furthermore, these problems stem from the inhomogeneity of the units with respect to the values of $R_c(u)$ and $Y_{c0}(u)$. This inhomogeneity is, I believe, the proper way to view the "disturbance terms" of the structural equations model (9) and (10) in section 3 (see section 4.4). One nice thing is that while the *direct effects* are not the same in Figures 5 and 4, the total effects are: Both equal $\tau + \rho\beta$.

4.3. *Two Different Ways to Estimate the Causal Effect of the Encouraged Activity*

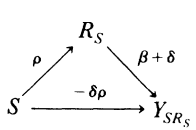
The message of the previous subsection is that the effect of study on test performance cannot be estimated by the usual regression methods of path analysis without making untestable assumptions about the counterfactual regression function, $\mu_c(r)$. If we assume that $\mu_c(r)$ is constant, then the biases shown in Figure 5 vanish and the usual path coefficients may be interpreted as causal effects, i.e. as ACEs. However, because $\mu_c(r)$ is so difficult to think about, there is little reason to believe that it is constant. Nor can δ be easily assessed as either positive or negative, since, in this example, there are reasons why it might be either: If students who study a lot tend to be those who do well even when they don't study, then δ is positive; but if those who study a lot are those who need to study, then δ is negative.

An alternative approach is to suppose that encouragement, of and by itself, has no effect on Y. In the studying example, this might be a plausible assumption. This corresponds to the restriction that

$$\tau = 0. \tag{55}$$

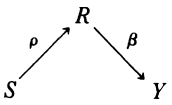
Now the empirical path diagram becomes that in Figure 6, and the path diagram for the causal theory becomes that in Figure 7. The total effect of S on Y_{SR_S} is now $\rho\beta$, whereas the total effect of S on R_S is ρ.

FIGURE 6.



Estimating beta (tau = 0 assume)

FIGURE 7.



S==G
IV for R

Hence,

$$\beta = \frac{\rho\beta}{\rho} = \frac{\text{total effect of } S \text{ on } Y_{SR_S}}{\text{total effect of } S \text{ on } R_S}. \tag{56}$$

This is also easily seen from the definitions of the ACEs and the FACEs. Under the assumption that $\tau = 0$,

$$\text{ACE}_{tc}(Y) = \text{FACE}_{tc}(Y) = \beta\rho, \tag{57}$$

regardless of whether or not $\mu_c(r)$ is linear, and hence

$$\beta = \frac{\text{ACE}_{tc}(Y)}{\text{ACE}_{tc}(R)} = \frac{\text{FACE}_{tc}(Y)}{\text{FACE}_{tc}(R)}. \tag{58}$$

The two FACEs in (58) may be estimated simply by the treatment-control mean difference in Y_{SR_S} and R_S, as mentioned earlier, so that (58) provides an alternative way to estimate β that does not assume that $\mu_c(r)$ is constant. In Powers and Swinton (1984), (58) was used to estimate β.

4.4. *Deriving a Structural Equations Model*

The ALICE model may be used to derive the structural equations model given in (9) and (10). If we substitute $S(u)$ for s in (40) and $S(u)$ for s and $R_S(u)$ for r in (41), we get the following pair of equations that involve the observables, S, R_S, Y_{SR_S}:

$$R_S(u) = R_c(u) + \rho S(u) \tag{59}$$

and

$$Y_{SR_S}(u) = Y_{c0}(u) + \tau S(u) + \beta R_S(u). \tag{60}$$

Now let

$$\eta_1(u) = R_c(u) - E(R_c)$$

and

$$\eta_2(u) = Y_{c0}(u) - E(Y_{c0}),$$

[redacted]

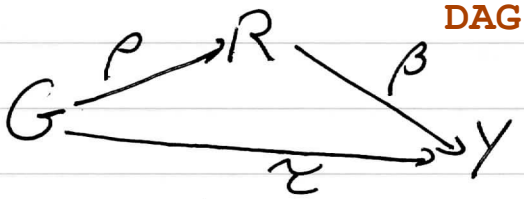
STAT 209 [redacted]

Encouragement Design Estimation (Holland 1988, p471)

Setting
Salt, BP
random assignment
to G=1 workshops, nagging
G=0 none
observe R salt intake
Y BP

Individual Picture (Holland ALICE)

DAG



Individual Potential outcomes

$R_G(u) = R_c(u) + \rho G$

$Y_{Gr}(u) = Y_{c0}(u) + \tau G + \beta r$

Regression Equations (over persons)

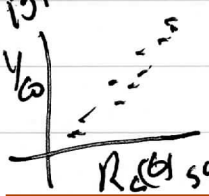
$R = \alpha_R + \gamma_1 G + \varepsilon_R \; ; \; Y = \alpha_y + \gamma_2 R + \gamma_3 G + \varepsilon_y$

Results for regression coefficients (Holland 1988)

$\gamma_1 = \rho \qquad \gamma_2 = \beta + \delta \qquad \gamma_3 = \tau - \rho\delta$

bias $\delta > 0$ association → salt intake when not encouraged

other healthy habits, attitudes → BP outcome when not encouraged and no reduction in salt



IF $\tau \equiv 0$ then algebra says use regression coefficients to solve

$\beta = \gamma_2 + \gamma_3/\gamma_1 = \beta + \delta - \delta = \beta$

Instrumental Variables IV (Week 6) requires $\tau = 0$ (basic IV assumption) G is instrument for R

$$\hat{\beta} = \frac{\hat{\beta}_{YG}}{\hat{\beta}_{RG}} = \frac{\bar{Y}_t - \bar{Y}_c}{\bar{R}_t - \bar{R}_c}$$ Wald estimator

note: tau = 0 is 'complete' mediation

Link between IV (Wald) and Holland estimator (tau = 0)

Barron-Kenney est $\hat{\gamma}_1 \hat{\gamma}_3$ path analysis indirect effect.

for $\tau = 0$

$$\beta = \gamma_2 + \gamma_3/\gamma_1 = \beta_{YR \cdot G} + \beta_{YG \cdot R}/\beta_{RG}$$

weak

Regression recursion

$$\beta_{YG \cdot R} = \beta_{YG} - \beta_{RG}\beta_{YR \cdot G}$$

Substitute

$$\gamma_2 + \gamma_3/\gamma_1 = \beta_{YR \cdot G} + \frac{\beta_{YG} - \beta_{RG}\beta_{YR \cdot G}}{\beta_{RG}}$$

$$= \beta_{YG}/\beta_{RG}$$

estimate by $\dfrac{\bar{Y}_1 - \bar{Y}_0}{\bar{R}_1 - \bar{R}_0}$ Wald estimator

Sesame Street IV

The Sesame Street example is from Andrew Gelman at Columbia,

Data from an experiment in which a randomly selected group of children
was encouraged to watch the television program Sesame Street and the randomly
selected control group was not. The goal of the experiment was to estimate the
effect on child cognitive development of watching more Sesame Street. In the
experiment, encouragement, but not actual watching, was randomized.

We might consider implementing a randomized experiment where the participants are
preschool children, the treatment of interest is watching Sesame Street, the control
condition is not watching, and the outcome is the score on a test of letter
recognition. It is not possible here for the experimenter to force children to watch
a TV show or to refrain from watching (the experiment took place while Sesame Street
was on the air). Thus watching cannot be randomized. Instead, when this study was
actually performed, what was randomized was encouragement to watch the show--this is
called a randomized encouragement design.

```
-------------------------
Code book with variable names for Sesame Street data
id : subject identification number
site :   1 =Three to five year old disadvantaged children from inner city areas in
             various parts of the country.
           2 = Four year old advantaged suburban children.
           3 = Advantaged rural children.
           4 = Disadvantaged rural children.
           5 = Disadvantaged Spanish speaking children.
sex   male=1, female=2
age   age in months
viewcat  frequency of viewing
             1=rarely watched the show
             2=once or twice a week
             3=three to five times a week
             4=watched the show on average more than 5 times a week
setting:    setting in which Sesame Street was viewed, 1=home 2=school
viewenc :  treatment condition
            1=child encouraged to watch,  2=child not encouraged to watch
encour:  treatment condition
           0=child not encouraged to watch, 1=child encouraged to watch
regular:  frequency of viewing:
          0=rarely watched the show, 1= watched once/week or greater
prebody :  pretest on knowledge of body parts (scores range from 0-32)
prelet :  pretest on letters (scores range from 0-58)
preform : pretest on forms (scores range from 0-20)
prenumb : pretest on numbers (scores range from 0-54)
prerelat : pretest on relational terms (scores range from 0-17)
preclasf : pretest on classification skills
postbody : posttest on knowledge of body parts (0-32)
postlet :  posttest on letters (0-58)
postform :  posttest on forms (0-20)
postnumb : posttest on numbers (0-54)
postrelat : posttest on relational terms (0-17)
postclasf:  posttest on classification skills
peabody:  mental age score obtained from administration of the Peabody Picture
           Vocabulary test as a pretest measure of vocabulary maturity
{note: measures used in analyses below are:
   encour, regular, postlet, prelet, peabody, site)
```

I have a data frame called "ses" ('cuz I can't spell sesame consistently)

```
> library(foreign)
> ses = read.dta("http://www.stat.columbia.edu/~gelman/arm/examples/sesame/sesame.dta")
> names(ses)
 [1] "rownames"  "id"        "site"      "sex"       "age"       "viewcat"   "setting"   "viewenc"
[18] "postnumb"  "postrelat" "postclasf" "peabody"   "agecat"    "encour"    "_Isite_2"  "_Isite_3"
> attach(ses)
> detach(ses)
> ses$imp = ses$postlet- ses$prelet
> attach(ses)

> dim(ses) #we have 240 children
```

```
[1] 240  28
The data frame is in wide form (each child is a row)
> names(ses) #approximately correspond to the codebook listing
 [1] "rownames"  "id"        "site"      "sex"     "age"    "viewcat"   "setting"
     "viewenc"   "prebody"   "prelet"
[11] "preform"   "prenumb"   "prerelat"  "preclasf"  "postbody"  "postlet"
       "postform"    "postnumb"  "postrelat" "postclasf"
[21] "peabody"   "agecat"    "encour"    "_Isite_2"  "_Isite_3"  "_Isite_4"
      "_Isite_5"  "regular"
```

Each of the 5 sites has encouraged and not encouraged (randomized) conditions
with different populations of children

```
> table(encour, site)
      site
encour  1  2  3  4  5
     0 28 19 14 23  4
     1 32 36 50 20 14
```

Effect of viewing Sesame Street on improvement in 'letters'.

```
> t.test(imp ~ encour)
        Welch Two Sample t-test
data:  imp by encour
t = -3.102, df = 173.592, p-value = 0.002244
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:   -7.567744 -1.682256
sample estimates:
mean in group 0 mean in group 1
          7.875          12.500
```

we have have information on actual viewing of Sesame Street by the children; here we will
use the dichotomous measure "regular". As stated in the introduction: "The goal of the experiment was
to estimate the effect on child cognitive development of watching more Sesame Street."

```
# useful descriptives
> tapply(postlet, encour, mean)
       0        1
24.92045 27.79605
> tapply(prelet, encour, mean)
       0        1
17.04545 15.29605
> tapply(prelet, regular, mean)
       0        1
14.42593 16.37634
> tapply(postlet, regular, mean)
       0        1
16.90741 29.59677

> tapply(postlet, list(encour,regular), mean)
         0        1
0 16.85000 31.64583
1 17.07143 28.88406
> tapply(prelet, list(encour,regular), mean)
         0        1
0 14.07500 19.52083
1 15.42857 15.28261

> tapply(regular, encour, mean)
        0         1
0.5454545 0.9078947
```

. Using the encouragement design formulation estimate of the effect on child cognitive development (
here) of watching more Sesame Street. What assumption is necessary for the IV estimation
in this design?
From the "useful descriptives" given you can reproduce this instrumental variables
estimate (Wald estimator).

```
> library(AER)
> ivimp = ivreg(imp ~ regular | encour)
> summary(ivimp)
Call:
```

```
ivreg(formula = imp ~ regular | encour)
Residuals:
    Min      1Q  Median      3Q     Max
-35.675  -7.675  -1.675   7.085  27.325
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9146     3.0185   0.303 0.762152
regular      12.7607     3.7995   3.359 0.000912 ***
---
Residual standard error: 10.28 on 238 degrees of freedom
Multiple R-Squared: 0.1562,     Adjusted R-squared: 0.1526
Wald test: 11.28 on 1 and 238 DF,  p-value: 0.0009123
> confint(ivimp)
                2.5 %    97.5 %
(Intercept) -5.001465  6.830673
regular      5.313858 20.207594
---------------------------------------
## Wald estimator
> tapply(imp, encour, mean)
     0      1
 7.875 12.500
> tapply(regular, encour, mean)
        0         1
0.5454545 0.9078947

> (12.5 - 7.875)/(.9079 - .5455)
[1] 12.76214
```

David A. Kenny
September 12, 2008

see handout pdf p39

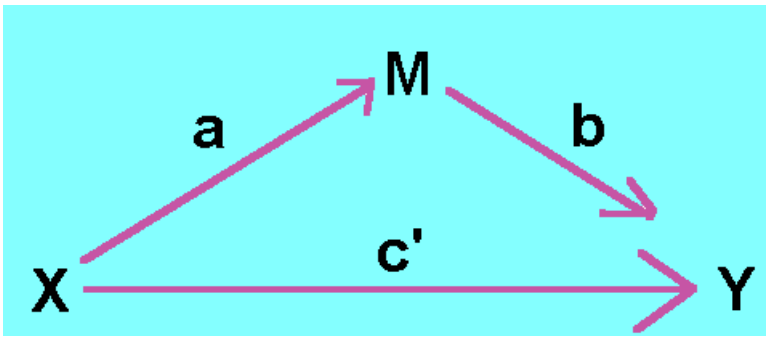
Learn how you can do a mediation analysis and output a text description of your results: Go to DataToText.

MEDIATION

What Is Mediation?

Consider a variable X that is assumed to affect another variable Y. The variable X is called the *initial variable* and the variable that it causes or Y is called the *outcome*. In diagrammatic form, the unmediated model is

The effect of X on Y may be mediated by a process or mediating variable M, and the variable X may still affect Y. Path c is called the *total effect*. The mediated model is



(These two diagrams are essential to the understanding of this page. Please study them carefully!) Path c' is called the *direct effect*. The mediator has been called an *intervening* or *process* variable. Complete mediation is the case in which variable X no longer affects Y after M has been controlled and so path c' is zero. Partial mediation is the case in which the path from X to Y is reduced in absolute size but is still different from zero when the mediator is controlled.

Note that a mediational model is a causal model. For example, the mediator is presumed to cause the outcome and not vice versa. If the presumed model is not correct, the results from the mediational analysis are of little value. Mediation is not defined statistically; rather statistics can be used to evaluate a presumed mediational model. The reader should consult the section below on Specification Error.

There is a long history in the study of mediation (Hyman, 1955; MacCorquodale & Meehl, 1948). Currently mediation is a very popular topic. (This page averages over 100 different visitors a day.) There are several reasons for the intense interest in this topic. One reason for testing mediation is trying to understand the mechanism through which the initial variable affects the outcome. Mediation (and moderation) analysis are a key part of what has been called ***process analysis.*** Moreover when most causal or structural models are examined, the mediational part of the model is the most interesting.

Baron and Kenny Steps

If the mediational model (see above) is correctly specified, the paths (c, a, b, and c') can be estimated by multiple regression, sometimes call ordinary least squares or OLS. As discussed later, other methods of estimation (e.g., logistic regression and structural equal modeling) can be used. Regardless of which data analytic method (the general assumption on this page is that the method is multiple regression) is used, the steps necessary for testing mediation are the same. This section describes the analyses required for testing mediational hypotheses [previously presented by Baron and Kenny (1986) and Judd and Kenny (1981)].

Baron and Kenny (1986) and Judd and Kenny (1981) have discussed four steps in establishing mediation:

A third reason for interest in mediation is methodological. Mediation represents the consideration of how a third variable affects the relation between two other variables. Although the consideration of a third variable may appear simple, three-variable systems can be very complicated, and there are many alternative explanations of observed relations other than mediation. This methodological and statistical challenge of investigating mediation has made methodology for assessing mediation an active research topic.

This review first defines the mediating variable and the ways in which it differs from other variables, such as a moderator or a confounder. Examples of mediating variables used in psychology are provided. Statistical methods to assess mediation in the single-mediator case are described, along with their assumptions. These assumptions are addressed in sections describing current research on the statistical testing of mediated effects, longitudinal mediation models, models with moderators as well as mediators, and causal inference for mediation models. Finally, directions for future research are outlined.

Definitions

Most research focuses on relations between two variables, X and Y, and much has been written about two-variable relations, including conditions under which X can be considered a possible cause of Y. These conditions include randomization of units to values of X and independence of units across and within values of X. Mediation in its simplest form represents the addition of a third variable to this X → Y relation, whereby X causes the mediator, M, and M causes Y, so X→ M → Y. Mediation is only one of several relations that may be present when a third variable, Z (using Z to represent the third variable), is included in the analysis of a two-variable system. One possibility is that Z causes both X and Y, so that ignoring Z leads to incorrect inference about the relation of X and Y; this would be an example of a confounding variable. In another situation, Z may be related to X and/or Y, so that information about Z improves prediction of Y by X, but does not substantially alter the relation of X to Y when Z is included in the analysis; this is an example of a covariate. Z may also modify the relation of X to Y such that the relation of X to Y differs at different values of Z; this is an example of a moderator or interaction effect. The distinction between a moderator and mediator has been an ongoing topic of research (Baron & Kenny 1986, Holmbeck 1997, Kraemer et al. 2001). A mediator is a variable that is in a causal sequence between two variables, whereas a moderator is not part of a causal sequence between the two variables. More detailed definitions of these variables in a three-variable system may be found in Robins & Greenland (1992).

Week 2

The single-mediator model is shown in **Figure 1**, where the variables X, M, and Y are in rectangles and the arrows represent relations among variables. **Figure 1** uses the notation most widely applied in psychology, with a representing the relation of X to M, b representing the relation of M to Y adjusted for X, and c' the relation of X to Y adjusted for M. The symbols e_2 and e_3 represent residuals in the M and Y variables, respectively. The equations and coefficients corresponding to **Figure 1** are discussed below. For now, note that there is a direct effect relating X to Y and a mediated effect by which X indirectly affects

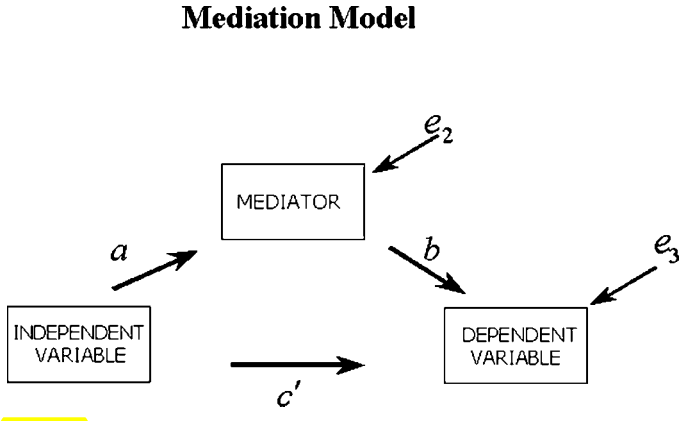


Figure 1
Mediation model.

Table 1 Subject area coverage in current mediation research

Subject area	# Articles cited
Social psychology	98
Clinical psychology	70
Health psychology	29
Developmental psychology	27
IO psychology	24
Cognitive psychology	18
Quantitative psychology (methods)	12
Program evaluation	8
Educational psychology	3
Environmental psychology	1
Evolutionary psychology	1

Y through M. Given that most prior mediation research has applied this single-mediator model, this review starts with this model. Limitations and extensions of the model are described in subsequent sections.

When thinking of mediation, it is helpful to understand that two models exist: One is theoretical, corresponding to unobservable relations among variables, and the other is empirical, corresponding to statistical analyses of actual data (MacCorquodale & Meehl 1948). The challenging task of research is to infer the true state of mediation from observations. There are qualifications even to this simple dichotomy, and in general, it will take a program of research to justify concluding that a third variable is a mediating variable.

Mediation in Psychological Research

In order to ascertain how often mediation is used in psychology, a search was conducted using the *PsycInfo* search engine for articles containing the word "mediation" in the title and citing the most widely cited article for mediation methods, Baron & Kenny (1986). This search yielded 291 references. Of these articles, 80 came from American Psychological Association (APA) journals. Publications earlier than the year 2000 were primarily APA sources, but there was a surge in non-APA articles after that time. The majority of these sources (239 citations) examined mediation alone, and 52 investigated both mediation and moderation effects. These studies included a mix of cross-sectional and longitudinal data, and ordinary least squares regression and structural equation modeling were the primary analytic methods. The articles covered a wide range of substantive areas, including social psychology (98 articles) and clinical psychology (70); a complete breakdown is listed in **Table 1**.

Mediation studies, such as those discussed above, are of two general but overlapping types. One type consists of investigating how a particular effect occurs. These studies usually occur after an observed X → Y relation is found. This approach stems from the elaboration methodologies outlined by Lazarsfeld (1955) and Hyman (1955). In this framework, a third variable is added to the analysis of an X → Y relation in order to improve understanding of the relation or to determine if the relation is spurious. A mediating variable improves understanding of such a relation because it is part of the causal sequence of X → M → Y. For example, physical abuse in early childhood is associated with violence later in life. One explanation of this pattern is that children exposed to physical violence acquire deviant patterns of processing social information that lead to later violent behavior. Dodge et al. (1990) found evidence for this theoretical mediating process because social processing measures explained the relation between early childhood physical abuse and later aggressive behavior.

The second type of study uses theory regarding mediational processes to design experiments. Some of the best examples of this approach are found in the evaluation of treatment and prevention programs. In this research, an intervention is designed to change mediating variables that are hypothesized to be causally related to a dependent variable. If the hypothesized relations are correct, a

mediation: R Package for Causal Mediation Analysis

Dustin Tingley Harvard
Teppei Yamamoto MIT
Luke Keele Penn State
Kosuke Imai Princeton

Abstract

In this paper, we describe the R package **mediation** for conducting causal mediation analysis in applied empirical research. In many scientific disciplines, the goal of researchers is not only estimating causal effects of a treatment but also understanding the process in which the treatment causally affects the outcome. Causal mediation analysis is frequently used to assess potential causal mechanisms. The **mediation** package implements a comprehensive suite of statistical tools for conducting such an analysis. The package is organized into two distinct approaches. Using the model-based approach, researchers can estimate causal mediation effects and conduct sensitivity analysis under the standard research design. Furthermore, the design-based approach provides several analysis tools that are applicable under different experimental designs. This approach requires weaker assumptions than the model-based approach. Finally, we also implement a statistical method for dealing with multiple (causally dependent) mediators, which are often encountered in practice.

Keywords: causal mechanisms, mediation analysis, **mediation**, R.

1. Introduction

Scholars across a range of disciplines are increasingly interested in identifying causal mechanisms, going beyond the estimation of causal effects. Once they ascertain that certain variables causally affect the outcome, the next natural step is to understand how these variables exert their influence. The standard procedure for analyzing causal mechanisms in applied research is called *mediation analysis*, where a set of linear regression models are fitted and then the estimates of "mediation effects" are computed from the fitted models (e.g., Haavelmo 1943; Baron and Kenny 1986; Shadish, Cook, and Campbell 2001; MacKinnon 2008). In recent years, however, causal mechanisms have been studied within the modern framework of causal inference with an emphasis on the assumptions required for identification. This approach has highlighted limitations of earlier methods and pointed the way towards a more flexible estimation strategy. In addition, new research designs have been proposed for identifying causal mechanisms.

In this paper, we introduce a full featured R package, **mediation**, for studying causal mechanisms. The **mediation** package allows users to (1) investigate the role of causal mechanisms using different types of data and statistical models, (2) explore how results change as identification assumptions are relaxed, and (3) calculate quantities of interest under alternative research designs. We focus on the demonstration of the functionalities available through the **mediation** package. The statistical theory that underlies the procedures implemented in the

David A. Kenny
September 12, 2008

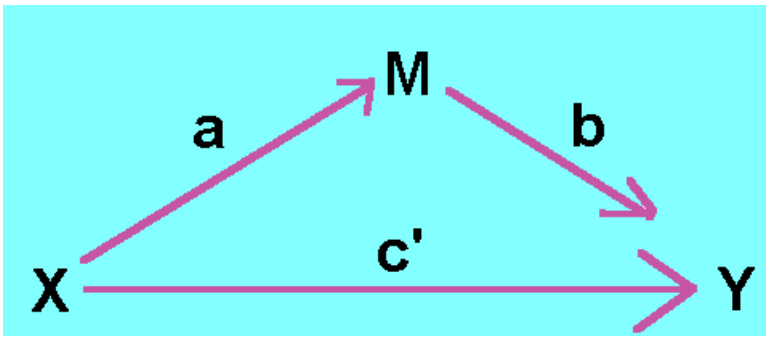
Learn how you can do a mediation analysis and output a text description of your results: Go to DataToText.

MEDIATION

What Is Mediation?

Consider a variable X that is assumed to affect another variable Y. The variable X is called the *initial variable* and the variable that it causes or Y is called the *outcome*. In diagrammatic form, the unmediated model is

The effect of X on Y may be mediated by a process or mediating variable M, and the variable X may still affect Y. Path c is called the *total effect*. The mediated model is



(These two diagrams are essential to the understanding of this page. Please study them carefully!) Path c' is called the *direct effect*. The mediator has been called an *intervening* or *process* variable. Complete mediation is the case in which variable X no longer affects Y after M has been controlled and so path c' is zero. Partial mediation is the case in which the path from X to Y is reduced in absolute size but is still different from zero when the mediator is controlled.

Note that a mediational model is a causal model. For example, the mediator is presumed to cause the outcome and not vice versa. If the presumed model is not correct, the results from the mediational analysis are of little value. Mediation is not defined statistically; rather statistics can be used to evaluate a presumed mediational model. The reader should consult the section below on Specification Error.

There is a long history in the study of mediation (Hyman, 1955; MacCorquodale & Meehl, 1948). Currently mediation is a very popular topic. (This page averages over 100 different visitors a day.) There are several reasons for the intense interest in this topic. One reason for testing mediation is trying to understand the mechanism through which the initial variable affects the outcome. Mediation (and moderation) analysis are a key part of what has been called ***process analysis.*** Moreover when most causal or structural models are examined, the mediational part of the model is the most interesting.

Baron and Kenny Steps

If the mediational model (see above) is correctly specified, the paths (c, a, b, and c') can be estimated by multiple regression, sometimes call ordinary least squares or OLS. As discussed later, other methods of estimation (e.g., logistic regression and structural equal modeling) can be used. Regardless of which data analytic method (the general assumption on this page is that the method is multiple regression) is used, the steps necessary for testing mediation are the same. This section describes the analyses required for testing mediational hypotheses [previously presented by Baron and Kenny (1986) and Judd and Kenny (1981)].

Baron and Kenny (1986) and Judd and Kenny (1981) have discussed four steps in establishing mediation:

Step 1: Show that the initial variable is correlated with the outcome. Use Y as the criterion variable in a regression equation and X as a predictor (estimate and test path c). This step establishes that there is an effect that may be mediated.

Step 2:
Show that the initial variable is correlated with the mediator. Use M as the criterion variable in the regression equation and X as a predictor (estimate and test path a). This step essentially involves treating the mediator as if it were an outcome variable.

Step 3:
Show that the mediator affects the outcome variable. Use Y as the criterion variable in a regression equation and X and M as predictors (estimate and test path b). It is not sufficient just to correlate the mediator with the outcome; the mediator and the outcome may be correlated because they are both caused by the initial variable X. Thus, the initial variable must be controlled in establishing the effect of the mediator on the outcome.

Step 4:
To establish that M completely mediates the X-Y relationship, the effect of X on Y controlling for M (path c') should be zero (see discussion on significance testing). The effects in both Steps 3 and 4 are estimated in the same equation.

If all four of these steps are met, then the data are consistent with the hypothesis that variable M completely mediates the X-Y relationship, and if the first three steps are met but the Step 4 is not, then partial mediation is indicated. Meeting these steps does not, however, conclusively establish that mediation has occurred because there are other (perhaps less plausible) models that are consistent with the data. Some of these models are considered later in the Specification Error section.

Note that the steps are stated in terms of zero and nonzero coefficients, not in terms of statistical significance, as they were in Baron and Kenny (1986). Because trivially small coefficients can be statistically significant with large sample sizes and very large coefficients can be nonsignificant with small sample sizes, the steps should not be defined in terms of statistical significance. Statistical significance is informative, but other information should be part of statistical decision making. For instance, consider the case in which a is large, b is zero, and so c = c'. It is very possible that the statistical test of c' is not significant (due to the collinearity of X and M) whereas c is significant. It would then appear that there is complete mediation when if fact there is no mediation at all.

Following, Kenny, Kashy, and Bolger (1998), one might ask whether all of the steps have to be met for there to be mediation. Certainly, Step 4 does not have to be met unless the expectation is for complete mediation. In the opinion of most though not all analysts, Step 1 is not required. However, note that a path from the initial variable to the outcome is implied if Steps 2 and 3 are met. If c' were opposite in sign to ab something that MacKinnon, Fairchild, and Fritz (2007) refer to as "inconsistent mediation," then it could be the case that Step 1 would not be met, but there is still mediation. In this case the mediator acts like a suppressor variable. Most analysts believe that the essential steps in establishing mediation are Steps 2 and 3.

James and Brett (1984) have argued that Step 3 should be modified by not controlling for the initial variable. Their rationale is that if there is complete mediation, there would be no need to control for the initial variable. However, because complete mediation does not always occur, it would seem sensible to control for X in Step 3.

Measuring Mediation or the Indirect Effect

The amount of mediation, which is called the *indirect effect*, is defined as the reduction of the effect of the initial variable on the outcome or c - c'. This difference in coefficients is theoretically exactly the same as the product of the effect of X on M times the effect of M on Y or ab; thus it holds that ab ≈ c - c'. The two are exactly equal when a) multiple regression (or structural equation modeling without latent variables) is used, b) there are no missing data, c) and the same covariates are in the equation. However, the two are only approximately equal for multilevel models, logistic analysis and structural equation model with latent variables. For such models, it is probably inadvisable to compute c from Step 1, but rather c (sometimes called the *total effect*, should be inferred to be c' + ab and not directly computed. Note that the amount of reduction in the effect of X on Y is not equivalent to either the change in variance explained or the change in an inferential statistic such as *F* or a *p* value. It is possible for the *F* from the initial variable to the outcome to decrease dramatically even when the mediator has no effect on the outcome! It is also not equivalent to a change in partial correlations.

If Step 2 (the test of a) and Step 3 (the test of b) are met, it follows that there necessarily is a reduction in the effect of X on Y. One way to test the null hypothesis that ab = 0 is to test that both a and b are zero (Steps 2 and 3). If such a strategy were used and one wanted a .05 probability of the combined test that a = 0 and b = 0, then alpha for the tests of a and b should lowered to .0253 so that the Type I error protection rate is correct.

It is much more common and more highly recommended (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) to perform a single test of ab is used . The test was first proposed by Sobel (1982). It requires the standard error of a or s_a (which equals a/t_a where t_a is the t test of coefficient a) and the standard error of b or s_b. The Sobel test provides the standard error of ab can be shown to equal approximately the square root of

$$b^2s_a^2 + a^2s_b^2$$

Other standard errors have been proposed, but the Sobel test has been by far the most commonly reported. (Though see below as bootstrapping seems to be replacing the more conservative Sobel test.) The test of the indirect effect is given by dividing ab by the square root of the above variance and treating the ratio as a Z test (i.e., larger than 1.96 in absolute value is significant at the .05 level). Kristopher J. Preacher and Geoffrey J. Leonardelli have an excellent web page that can help you calculate these test (go to the Sobel test). Measures and tests of indirect effects are also available within many structural equation modeling programs. These programs appear to use the Sobel formula.

The derivation of the Sobel standard error presumes that a and b are independent, something that is true when the tests are from multiple regression but not true when other tests are used (e.g., logistic regression, structural equation modeling, and multilevel modeling). In such cases, the researcher ideally provides evidence for approximate independence. Additionally, the Sobel test can be conducted using the standardized or unstandardized coefficients. Care must be taken to use the appropriate standard errors if standardized coefficients are used.

The Sobel test is very conservative (MacKinnon, Warsi, & Dwyer, 1995), An increasingly popular alternative is bootstrapping (Shrout & Bolger, 2002) which is beginning to replace the Sobel test of the indirect effect. Hayes and Preacher have written SPSS and SAS macros that can be downloaded for tests of indirect effects (click here to go Kris Preacher's page). Also Amos can be used to bootstrap (click here for a tutorial).

A related measure of mediation is the proportion of the effect that is mediated or 1 - ab/c. Such a measure while theoretically informative is very unstable and should not be computed is c is small. Note too that it is greater than 1 when there is inconsistent mediation. The measure can be informative, especially when c' is not statistically significant. See the example in Kenny et al. (1998) where c' is not statistically significant but only 56% of c is explained.

Mediation / Moderation

STAT 209 [redacted]

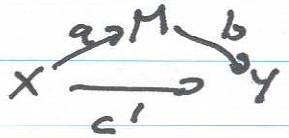
Sources

Kenny web page (Stat209 week 2)

MacKinnon Ann Rev Psych 2007 linked observational

Kraemer Arch Gen Psych linked RCT

Baron-Kenny $X \xrightarrow{c} Y$



OLS (or logistic) for a b c c'

steps $\rho_{XY} \neq 0$ (1) $\rho_{XM} \neq 0$ (2) $\beta_{YM \cdot X} \neq 0$ (3) (b) $\rho_{YX \cdot M} = 0$ (4) complete

Gauss-normal eq's

$$\beta_{12} = \beta_{12 \cdot 3} + \beta_{32}\beta_{13 \cdot 2}$$

(c) (c') (a) (b)

Y=1 X=2 M=3

$c - c' = \beta_{32}\beta_{13 \cdot 2} = ab$ amount of mediation

Inference "Sobel" $Var(c - c') = b^2 s_a^2 + a^2 s_b^2$

asymptotic var product 2 rv (Mood/Graybill)

delta-method (first term)

R-implementation Multilevel package (pdf)

Bound measurement error Cochrane (Stat 209 week 1)

Holland encouragement designs (Stat 209 week 2,3) next Mtg

MacKinnon Fig 1 Eq 1-3

plot pp 599-600

inf Sobel, bootstrap mediated moderation CNRL p 606

Kraemer Fig 1 table

moderator pre-existing mediator, during treatment.

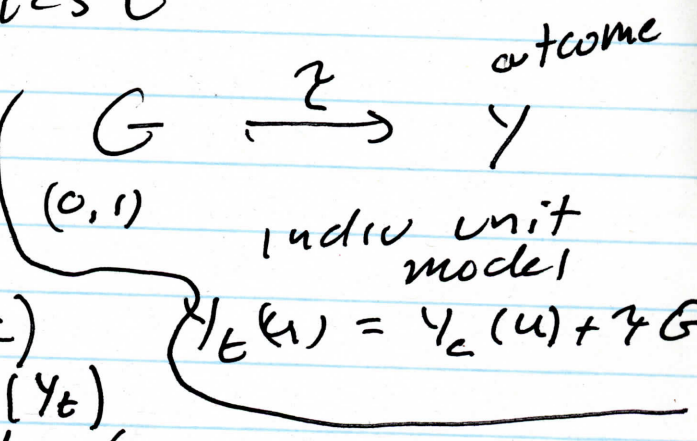
Stat 209 Week 2, Mediating variables numerical example

```
> dim(ssdat)
[1] 137  12
> summary(ssdat)
     grade           gender       schoolgrades
 Min.   :6.000   Min.   :1.000   Min.   :1.000
 1st Qu.:6.000   1st Qu.:1.000   1st Qu.:4.000
 Median :7.000   Median :2.000   Median :6.000
 Mean   :6.993   Mean   :1.526   Mean   :5.762
 3rd Qu.:8.000   3rd Qu.:2.000   3rd Qu.:8.000
 Max.   :8.000   Max.   :2.000   Max.   :9.000
                                 NA's   :7
     belong         depress          master
 Min.   :32.00   Min.   :10.00   Min.   : 7.00
 1st Qu.:52.00   1st Qu.:14.00   1st Qu.:40.00
 Median :61.00   Median :20.00   Median :54.00
 Mean   :61.77   Mean   :19.68   Mean   :50.79
 3rd Qu.:72.00   3rd Qu.:25.00   3rd Qu.:61.00
 Max.   :89.00   Max.   :37.00   Max.   :80.00
                 NA's   :1

> cor(ssdat, use = "pairwise.complete.obs")
             schoolgrades       belong     depress      master
schoolgrades   1.00000000   0.32461877 -0.23805195  0.25281286
belong         0.32461877   1.00000000 -0.50984362  0.59237433
depress       -0.23805195  -0.50984362  1.00000000 -0.46763766
master         0.25281286   0.59237433 -0.46763766  1.00000000

X = Belong; M = mastery; Y= Depress
Y on X regression; is M a mediator?
> reg1 = lm(depress ~ belong)
> summary(reg1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.59112    2.36527  15.047  < 2e-16 ***
belong      -0.25715    0.03748  -6.861  2.3e-10 ***
Residual standard error: 5.386 on 134 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared: 0.2599,     Adjusted R-squared: 0.2544

> library(multilevel)
> ?sobel
> sobel(belong, master, depress)
$`Mod1: Y~X`
              Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 35.5911156 2.36526995 15.047380 1.400184e-30
pred        -0.2571453 0.03748198 -6.860505 2.303242e-10

$`Mod2: Y~X+M`
              Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 36.4998362 2.32596622 15.692333 4.665379e-32
pred        -0.1810422 0.04516544 -4.008423 1.013040e-04
med         -0.1103938 0.03855989 -2.862917 4.878432e-03

$`Mod3: M~X`
             Estimate Std. Error  t value     Pr(>|t|)
(Intercept) 8.2316307  5.1621818 1.594603 1.131570e-01
pred        0.6893784  0.0818041 8.427187 4.875161e-14
```

```
$Indirect.Effect     [1] -0.07610307
$SE       [1] 0.02807444
$z.value  [1] -2.71076
$N  [1] 136
> #master was a better choice than schoolgrades, it shows up
-----------------------------------------------------------------------
> install.packages("MBESS")
> library(MBESS)

> mediation(belong, master, depress)
$Y.on.X
$Y.on.X$Regression.Table
                Estimate Std. Error   t value      p(>|t|) Low Conf Limit Up Conf Limit
Intercept.Y_X         NA 2.36100648        NA           NA             NA            NA
c (Regressor) -0.2571453 0.03748198 -6.860505 2.303242e-10     -0.3312781    -0.1830125

$M.on.X
$M.on.X$Regression.Table
               Estimate Std. Error  t value      p(>|t|) Low Conf Limit Up Conf Limit
Intercept.M_X 8.2028494  5.1528768 1.591897 1.137642e-01     -1.9886431    18.3943419
a (Regressor) 0.6893784  0.0818041 8.427187 4.875161e-14      0.5275842     0.8511727

$Y.on.X.and.M
$Y.on.X.and.M$Regression.Table
                     Estimate Std. Error   t value     p(>|t|) Low Conf Limit Up Conf
Intercept.Y_XM             NA 2.32170030        NA          NA             NA
c.prime (Regressor) -0.1810422 0.04516544 -4.008423 0.000101304     -0.2703777   -0.091
b (Mediator)        -0.1103938 0.03855989 -2.862917 0.004878432     -0.1866637   -0.034

$Effect.Sizes
                                                  Estimate
Indirect.Effect                                -0.07610307
$Bootstrap.Results
                      Estimate CI.Lower_Percentile   CI.Upper_Percentile
Indirect.Effect     -0.07610307        -0.135213084           -0.02573436

                                CI.Lower_BCa    CI.Upper_BCa
                                -0.135215775    -0.025764195
```

Potential outcomes (recap)
initial designs

t-test, bivariate t-test



random assignment

$E(Y_t \mid S=t) = E(Y_t \mid S=c) = E(Y_t)$

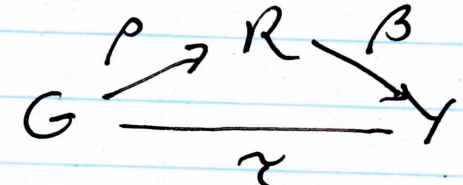
otherwise bias in t-test

Extended Designs
mediation, encouragement

even if G represents random assignment can only estimate ρ easily.

Estimation of β or $\rho * \beta$

encour mediation



indiv unit model from handouts

more difficult. Path analysis fails (bias δ)

IV assumption $\tau = 0$

$\Rightarrow$ total mediation (prob 5, HW 3)

Mediator Model Types	*Outcome Model Types*						
	Linear	GLM	Ordered	Censored	Quantile	GAM	Survival
Linear (`lm`)	✓	✓	✓*	✓	✓	✓*	✓
GLM (`glm`/`bayesglm`)	✓	✓	✓*	✓	✓	✓*	✓
Ordered (`polr`/`bayespolr`)	✓	✓	✓*	✓	✓	✓*	✓
Censored (tobit via `vglm`)	-	-	-	-	-	-	-
Quantile (`rq`)	✓*	✓*	✓*	✓*	✓*	✓*	✓
GAM (`gam`)	✓*	✓*	✓*	✓*	✓*	✓*	✓*
Survival (`survreg`)	✓	✓	✓*	✓	✓	✓*	✓

Table 1: Types of Statistical Models That Can be Used with the `mediate` Function. Asterisks, *, indicate the model combinations that can only be estimated using the nonparametric bootstrap (i.e. with the argument `boot = TRUE` for the `mediate` function).

this as a special case and accommodates a greater range of statistical models as shown in Table 1.

We now illustrate the use of the `mediate` function with an empirical example based on the `framing` data of Brader, Valentino, and Suhat (2008). This data set is a part of the **mediation** library and can be loaded via the following syntax,

```
> ## load the package
> library(mediation)
> ## load the framing data
> data(framing)
```

A brief description of each variable in the data can be obtained through a help file,

```
> ?framing
```

Brader et al. (2008) conducted a randomized experiment where subjects are exposed to different media stories about immigration and the authors investigated how their framing influences attitudes and political behavior regarding immigration policy. They posit the role of anxiety as the mediating variable for the causal effect of framing on public opinion. We first fit the mediator model where the measure of anxiety (`emo`) is modeled as a function of the framing treatment (`treat`) and pretreatment covariates (`age`, `educ`, `gender`, and `income`). Next, we model the outcome variable, which is a binary variable indicating whether or not the participant agreed to send a letter about immigration policy to his or her member of Congress (`cong_mesg`). The explanatory variables of the outcome model include the mediator, treatment status, and the same set of pre-treatment variables as those used in the mediator model. In this example, the treatment is expected to increase the level of respondents' emotional response, which in turn is hypothesized to make subjects more likely to send a letter to his or her member of Congress. We use the linear least squares regression and the probit regression for the mediator and outcome models, respectively.

```
> ## Mediator Model
> med.fit <- lm(emo ~ treat + age + educ + gender + income, data = framing)
```

that only the treatment is randomized. Unfortunately, Imai et al. (2012) derive the bounds in the case with all binary variables (treatment, mediator, and outcome) and show that the bounds are never informative about the sign of the ACME (i.e., they always cover 0).

Most mediation analysis proceeds under the sequential ignobility assumption. Those analyses also tend to be model-based, but they need not be. Imai et al. (2010c) outline a design-based estimator for the ACME for when the mediator is discrete. This estimator for the ACME is fully nonparametric. One drawback to this estimator is that one can encounter mediator-treatment combinations for which there are no subjects because of data sparsity. Standard error calculation for this estimator is based on either the Delta method or the nonparametric bootstrap.

The `mediate.sed` function requires the names of the outcome, mediator, and treatment variables, along with the name of the data frame that contains these variables. When `SI = TRUE`, the function will return the point estimates under the sequential ignorability assumption, and otherwise the results will be a set of sharp bounds for the ACME. The method for inference also differs slightly from the `mediate` function. When `boot = TRUE` the bootstrap is used, but when `boot = FALSE`, the Delta method is used to compute standard errors.

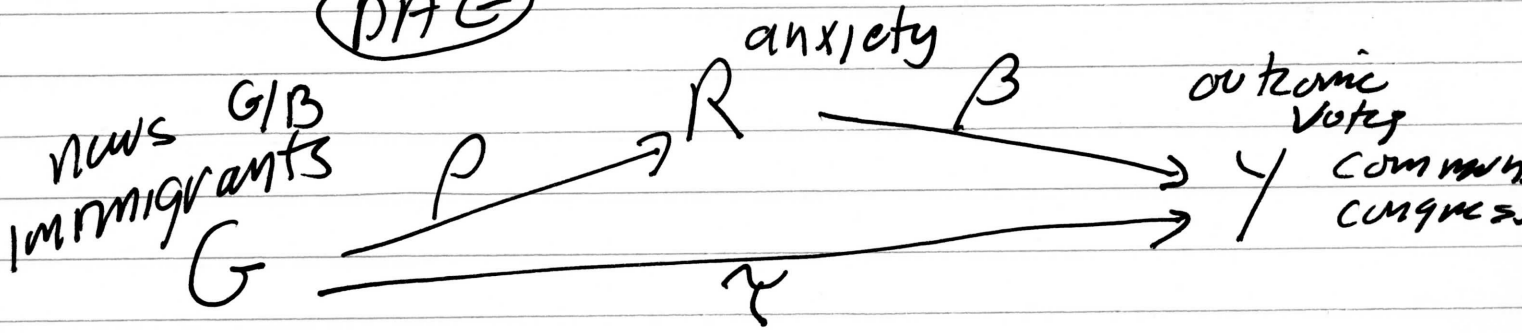
Below, we present an example using the framing data from Brader et al. (2008). The treatment variable is the same as before, i.e., `treat`, and the mediator is `anx`, which refers to a subject's reported level of anxiety. This four level measure is one component of the `emo` variable that was previously used as the mediator and in the data all treatment-mediator combinations are present (a requirement for the estimator). The outcome variable in this example is `english` and measures on a four point scale how much someone supports English only laws, from strongly support to strongly oppose. Note that the `mediate.sed` function only take numeric variables as arguments. Variables that are stored as factors must be converted to numeric variables as we show below.

```
> framing$english <- as.numeric(framing$english)
> framing$anx <- as.numeric(framing$anx)
> sed.est <- mediate.sed("english", "anx", "treat", data = framing, SI = TRUE,
+                        boot = TRUE)
> summary(sed.est)


Design-Based Causal Mediation Analysis

Single Experiment Design with Sequential Ignorability

Confidence Intervals Based on Nonparametric Bootstrap

Mediation Effect_0:  0.1021 95 % CI  -0.6378  0.8438
Mediation Effect_1:  0.07066 95 % CI  -0.2099  0.3439
Sample Size Used:    265
```

The results from the `summary` function display the mediation effects along with the default 95% confidence intervals. In this example both $\bar{\delta}(0)$ and $\bar{\delta}(1)$ are not significantly different from 0.

Framing (mediation)

PATE



path analysis

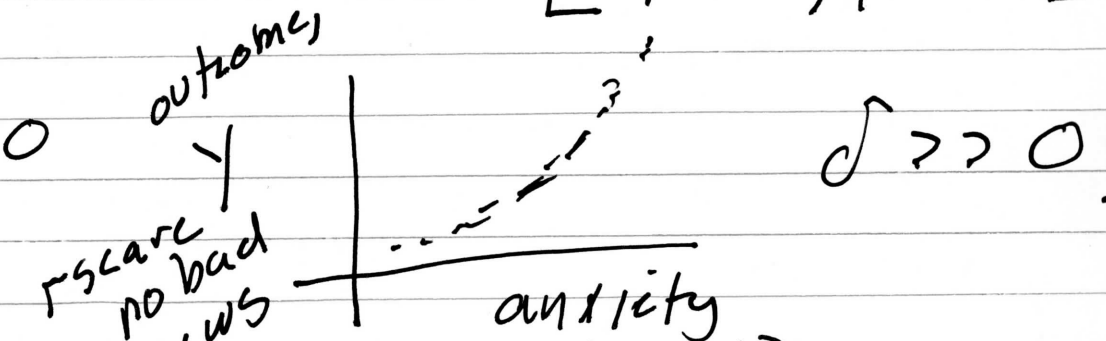
$R \sim G$ (ρ?)

$Y \sim G + R$ (τ?, β?)

(money)

$[\tau - \rho\delta, \beta + \delta]$

here $\delta \neq 0$



$\delta >> 0$

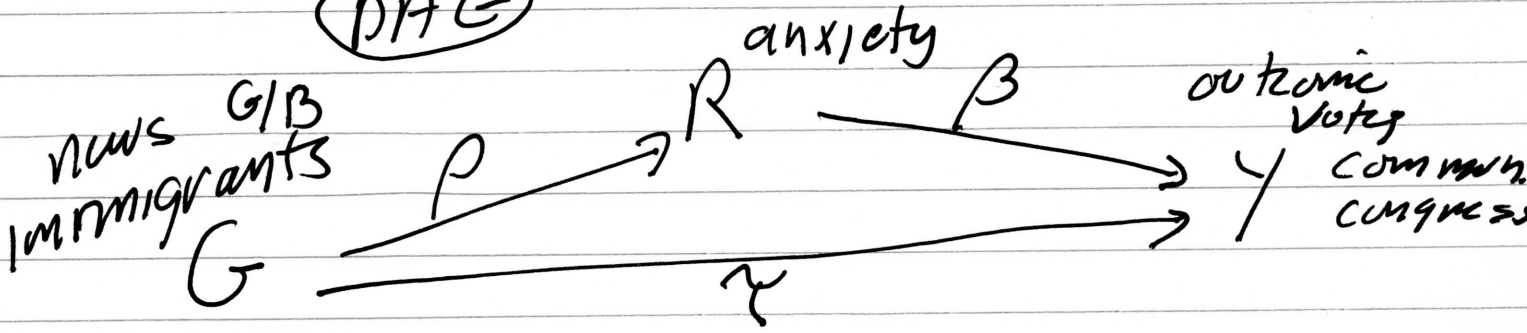
mediation effect is $\rho\beta$

estimate that? only for total mediation.

[lots of use in observational studies--can't get rho, total effect.

Framing (mediation)

PATE



scare/not scare

path analysis

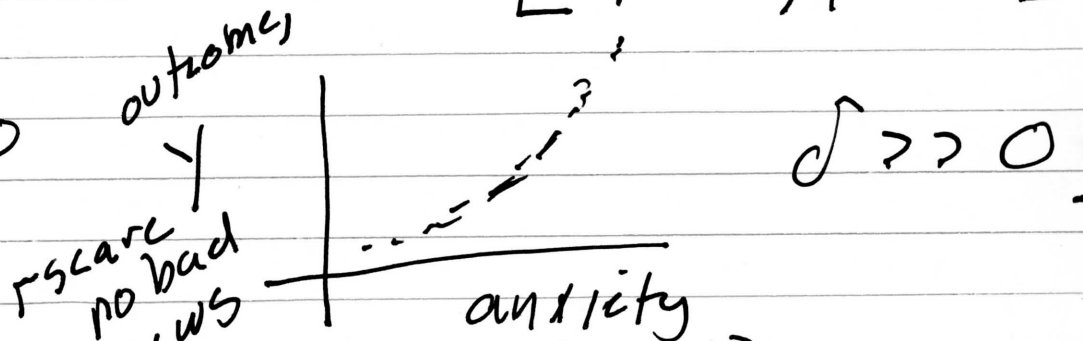
$R \sim G$ (ρ?)

$Y \sim G + R$ (τ?, β?)

(money)

$[\tau - \rho\delta, \beta + \delta]$

here $\delta \neq 0$



$\delta >> 0$

mediation effect is $\rho\beta$

estimate that? only for total mediation.

The Red Sneakers Effect: Inferring Status and Competence from Signals of Nonconformity

SILVIA BELLEZZA
FRANCESCA GINO
ANAT KEINAN

This research examines how people react to nonconforming behaviors, such as entering a luxury boutique wearing gym clothes rather than an elegant outfit or wearing red sneakers in a professional setting. Nonconforming behaviors, as costly and visible signals, can act as a particular form of conspicuous consumption and lead to positive inferences of status and competence in the eyes of others. A series of studies demonstrates that people confer higher status and competence to nonconforming rather than conforming individuals. These positive inferences derived from signals of nonconformity are mediated by perceived autonomy and moderated by individual differences in need for uniqueness in the observers. An investigation of boundary conditions demonstrates that the positive inferences disappear when the observer is unfamiliar with the environment, when the nonconforming behavior is depicted as unintentional, and in the absence of expected norms and shared standards of formal conduct.

> Your sweats, PJs and flip-flops are losing you money! . . . Do you crave more confidence, respect and power? . . . Find out how image connects to success. (Eve Michaels, author of *Dress Code*)
>
> I have a number of super-successful Silicon Valley clients who dress in ripped denim, Vans shoes, and T-shirts. They are worth hundreds of millions, even more, but it's a status symbol to dress like you're homeless to attend board meetings. (Tom Searcy, *CBS Moneywatch*)

In both professional and nonprofessional settings, individuals often make a significant effort to learn and adhere to dress codes, etiquette, and other written and unwritten standards of behavior. Conformity to such rules and social norms is driven by a desire to gain social acceptance and status (see Cialdini and Goldstein 2004) and avoid negative sanctions such as social disapproval, ridicule, and exclusion (Kruglanski and Webster 1991; Levine 1989; Miller and Anderson 1979; Schachter 1951). In the present research, we propose that under certain conditions, nonconforming behaviors can be more beneficial than efforts to conform and can signal higher status and competence to others. We argue that while unintentional violations of normative codes and etiquette can indeed result in negative inferences and attributions, when the deviant behavior appears to be deliberate, it can lead to higher rather than lower status and competence inferences.

Since nonconformity often has a social cost (e.g., Levine 1989; Schachter 1951), observers may infer that a nonconforming individual is in a powerful position that allows her to risk the social costs of nonconformity without fear of losing her place in the social hierarchy. Signaling theory suggests that, for a signal to be effective, it must be costly and observable by others (Feltovich, Harbaugh, and To 2002; Spence 1973; Zahavi and Zahavi 1997). We propose that nonconforming behaviors, as costly and observable signals, can act

Silvia Bellezza (sbellezza@hbs.edu) is a doctoral candidate in marketing, Francesca Gino (fgino@hbs.edu) is associate professor of business administration, and Anat Keinan (akeinan@hbs.edu) is associate professor of marketing at the Harvard Business School, Boston, MA 02163. The authors are grateful for helpful comments received from participants in seminars and lab groups at Harvard and MIT. They thank John T. Gourville, Max H. Bazerman, Zoe Chance, Michael I. Norton, and Natalie Zakarian for their constructive feedback on earlier drafts of the article. The article is based on part of the first author's dissertation.

Mary Frances Luce served as editor and Rebecca Ratner served as associate editor for this article.

Electronically published December 18, 2013

If you're a billionaire in Silicon Valley, you can wear what you want on your feet. Mark Zuckerberg, the C.E.O. of Facebook, has made numerous public appearances in a hoodie and Adidas slide-on sandals. Sergey Brin, a co-founder of Google, is not shy about his Vibram FiveFingers barefoot-style athletic shoes. And Steve Jobs, the late C.E.O. of Apple, was rarely seen without gray New Balance running sneakers. "I have a number of super-successful Silicon Valley clients who dress in ripped denim, Vans shoes and T-shirts," the consultant Tom Searcy has written. "It's a status symbol to dress like you're homeless to attend board meetings."

While people generally adhere to group norms for fear of disapproval or reprimand, anecdotal evidence and the occasional study suggest that high-status folk feel free to break rules—by eating with their mouths open, violating traffic laws, and expressing unpopular opinions. But how is nonconformity interpreted by others? Do we see it as a sign of status? New research, to be published next near in *The Journal of Consumer Research*, suggests that we do. The authors call the phenomenon the "red sneakers effect," after one of them taught a class at Harvard Business School in her red Converse.

Silvia Bellezza, a doctoral candidate at Harvard Business School, and Francesca Gino and Anat Keinan, two professors there, first studied the link between accomplishment and informality. They found that scholars who dressed down at an academic conference, eschewing blazers for T-shirts, had stronger research records, even controlling for age and gender. Then, they explored why and when this sartorial tactic for announcing status—if that's what it is—succeeds.

Bellezza went to Milan and asked some clerks at luxury boutiques (Armani, Valentino, etc.) to imagine a woman entering the store wearing either gym clothes or a fur coat. Others were to imagine a woman in flip-flops and a Swatch, or in high heels and a Rolex. Clerks then judged her likely financial and celebrity status. Of the hypothetical shoppers, the casually attired were judged wealthier and more important. One clerk said, "Wealthy people sometimes dress very badly to demonstrate superiority," and that "if you dare enter these boutiques so underdressed, you are definitely going to buy something." But when Bellezza ran the same questions by local pedestrians, they assumed a done-up client to be wealthier. Picking up on status cues, the researchers determined, seems to require familiarity with the environment in which those cues are used.

Next, the researchers asked students at American universities to imagine a professor who is clean-shaven and wears a tie, or one who is bearded and wears T-shirts. Students were slightly more inclined to judge the dapper professor as a better teacher and

researcher. But some students were given another piece of information: that the professor works at a top-tier school, where the dress code is presumably more formal. For them, the slouchy scholar earned more points. Deviance can signal status, but only when there are clear norms from which to deviate.

What if you stand out not for informality but for originality? In another experiment, a hypothetical man wearing a red bow tie at a black-tie party hosted by his golf club was viewed as higher in status—and better at golf—than a peer who stuck with the black-tie dress code. But if subjects were told the man broke the dress code unintentionally, he gained no benefit. When it's not clear that a person is breaking a norm deliberately, he might be seen merely as missing the memo, or not having the wherewithal to follow it.

The next study looked for clues about why we see nonconformity as a sign of status. Subjects evaluated a hypothetical M.I.T. student presenting a business plan in a competition. He used the M.I.T. PowerPoint template others were using, or he used his own. As predicted, participants saw the one who abandoned the standard template as having a better business idea, and as being more respected by his friends. They also rated him as more autonomous—someone who "can afford to do what he wants." Further, people perceived the nonconformist as having high status and competence, because he seemed to act autonomously.

The red-sneaker effect fits in with a wider body of research on the idea that certain observable traits or behaviors signal hidden qualities by virtue of their "costliness." For instance, a peacock's colorful tail feathers make it easy prey for predators, but they tell a peahen that he's fit enough to sustain the risk. The more one has of the trait to be touted (fitness, say), the less costly the signal (feathers), making the display of the signal a reliable proxy for the trait. This is how conspicuous consumption works: jewelry is costly, unless you're rich and won't miss the cash. Similarly, deliberate nonconformity shows that you can handle some ridicule because you've got social capital to burn.

The economist Nick Feltovich and his colleagues have done work demonstrating that this kind of behavior—known as costly signalling—can also lead high-status people to avoid being ostentatious. Imagine three groups of people: those with low, medium, and high amounts of a desirable trait, like wealth. Someone without much income would have to make big sacrifices to buy a BMW. If you've got a bit more money—you're a medium—it's easier for you to signal wealth, and you might buy status symbols so that no one mistakes you for a poor person. A really wealthy person, on the other hand—a high—can distinguish himself from the mediums by choosing *not* to send costly signals of wealth. If he has enough secondary signals of status—a prime address, a high-profile list of friends—he'll feel secure in not being mistaken for poor. (Understatement can also work when signalling talent, popularity, or intellect. Thus, Harvard graduates say only that they went to school "in Boston.")

Other recent research demonstrates that when people violate norms, it makes them seem powerful. In one experiment reported in 2011, by Gerben Van Kleef of the University of Amsterdam, along with colleagues, subjects imagined a scene in a city-hall waiting room. Someone stands up and either goes to the bathroom or grabs coffee from a worker's desk. The coffee-taker was rated as more decisive and in control and as having more authority and influence. In another experiment, a bookkeeper who said it's O.K. to bend the rules was seen as having more power than one who said it wasn't. And in another, subjects rated a fellow subject as more powerful when he arrived late to the experiment and put his feet up on the table.

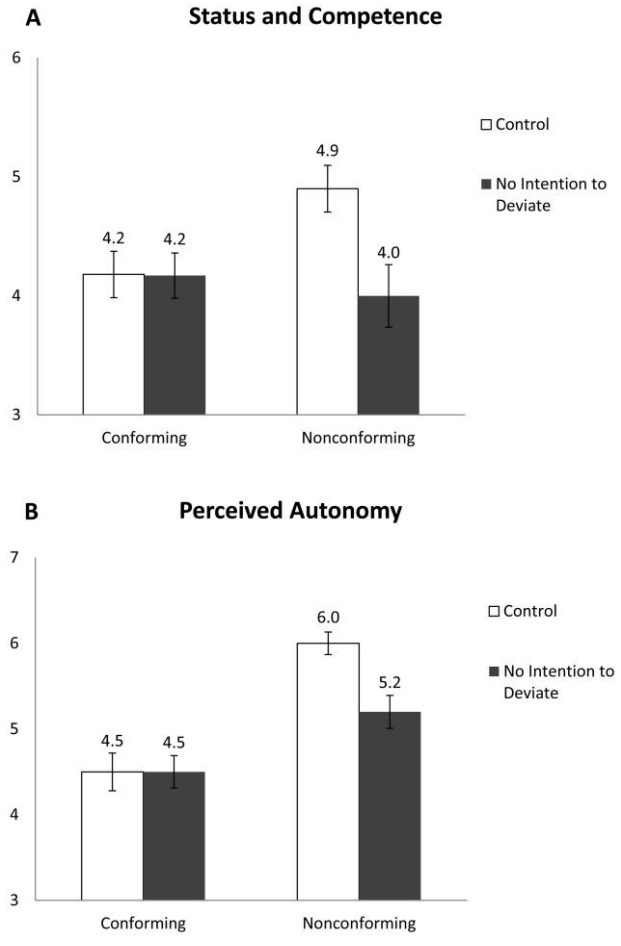
The researchers raise several questions about their findings, such as whether people in other countries and cultures would make the same inferences as their counterparts in the U.S. and Italy. Or whether there's a sweet spot for nonconformity: too much, and maybe you're a jerk or a weirdo. Or whether the attractiveness of the nonconformist matters. Or how one can signal that one is flouting a convention intentionally. But, for now, the conclusion seems to be that, in the right situation, breaking the rules a little can be a great way to show off—assuming you can back it up.

Gino, one of the co-authors of the Harvard research, recently received tenure at Harvard Business School; she avoids the school's PowerPoint templates and loves her red Converse. I asked Bellezza if she ever uses nonconformity as a signal. "I'm still a doctoral student," she said with a laugh. "I wouldn't dare teach a class in red sneakers myself. Maybe one day, but not now."

Matthew Hutson is a science writer in New York City and the author of "The 7 Laws of Magical Thinking: How Irrational Beliefs Keep Us Happy, Healthy, and Sane."

FIGURE 4

STUDY 3 RESULTS: UNINTENTIONALITY OF THE NONCONFORMING BEHAVIOR AS BOUNDARY CONDITION



($M_{\text{control nonconformity}} = 4.9$ vs. $M_{\text{control conformity}} = 4.2$, $t(68) = 2.4$, $p < .05$). In contrast, there was no significant difference between conditions when the behavior was depicted as unintentional ($M_{\text{unintentional nonconformity}} = 4.0$ vs. $M_{\text{unintentional conformity}} = 4.2$, NS). Importantly, when comparing the two nonconforming conditions, we found that the positive status and competence inferences associated with wearing a red bow tie significantly decreased when the nonconforming behavior was clearly depicted as unintentional ($M_{\text{control nonconformity}} = 4.9$ vs. $M_{\text{unintentional nonconformity}} = 4.0$, $t(68) = 2.6$, $p < .001$). We also checked whether participants' gender would impact status attributions. We conducted the same ANOVA analysis including gender as a control variable and found no significant effect for this demographic variable.

Perceived Autonomy. We then performed a similar analysis using ratings of autonomy as the dependent variable. The analysis revealed a significant main effect for nonconformity ($F(1, 137) = 35.3$, $p < .001$), a significant main effect for deliberateness ($F(1, 137) = 4.1$, $p < .05$), and a significant interaction ($F(1, 137) = 4.8$, $p < .05$), depicted in figure 4B. As predicted, when we provided no information about the intentions of the described individual and the behavior was interpreted as deliberate, participants perceived the nonconforming individual as having significantly higher autonomy than the conforming one ($M_{\text{control nonconformity}} = 6.0$ vs. $M_{\text{control conformity}} = 4.5$, $t(68) = 6.0$, $p < .001$). The nonconformity manipulation elicited a significant difference between conditions also when the behavior was depicted as unintentional ($M_{\text{unintentional nonconformity}} = 5.2$ vs. $M_{\text{unintentional conformity}} = 4.5$, $t(69) = 2.6$, $p < .05$). Importantly, the comparison between the two nonconforming conditions revealed that the perceived autonomy participants associated with wearing a red bow tie was significantly weakened when this nonconforming behavior was depicted as unintentional ($M_{\text{control nonconformity}} = 6.0$ vs. $M_{\text{unintentional nonconformity}} = 5.2$, $t(68) = 3.4$, $p < .001$), as we predicted. In sum, we find that enhanced perceptions of the nonconforming individual's status, competence, and autonomy dissipate when observers perceive the nonconforming conduct as unintentional.

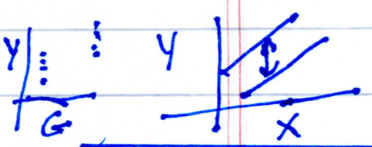
Mediated Moderation Analysis. To test moderation by deliberateness and mediation by perceived autonomy, we conducted a mediated moderation analysis (Edwards and Lambert 2007) examining whether perceived autonomy mediated the detected interaction between nonconformity and deliberateness. As reported above, deliberateness significantly moderated both the dependent variable (status and competence) and the mediator (autonomy). Moreover, when status and competence were regressed on nonconformity, deliberateness, their two-way interaction, and autonomy, the mediator was significant ($B = .38$, $t(137) = 4.1$, $p < .001$), and the effect of the interaction between nonconformity and deliberateness on status and competence became nonsignificant (from $B = .82$, $t(137) = 1.9$, $p = .05$, to $B = .51$, $t(136) = 1.2$, NS). In a bootstrap analysis, we found that the 95% bias-corrected confidence interval for the size of the indirect effect excluded zero (95% CI = .033 to .717), suggesting a significant indirect effect.

Taken together, the results of study 3 deepen our understanding of the interactions among the underlying processes of the red sneakers effect. We find that nonconformity leads to inferences of heightened status and competence, as long as the deviant conduct is perceived as deliberate. Moreover, we show that autonomy mediates the interaction between the nonconformity manipulation and deliberateness on status and competence inferences.

Discussion. Study 3 extends our previous findings by examining deviance from the norm through a dress choice that denotes originality. We demonstrate that participants perceive an individual wearing a red bow tie at a black-tie party in a country club as a higher-status member in the club and as a better golf player relative to a conforming individual wearing a black bow tie. Importantly, this study explores the role of perceived deliberateness as a boundary

Recap, RCT, G = T/C (0,1) Y outcome

Before ... ATE via t-test Y ~ G est $\mu_1 - \mu_0$
[stat 60] (ACE) or using X (concomitant var)
ancova (or blocking) Y ~ G + X
for precision: coef G est $\mu_1 - \mu_0$



But there's more ...

mediation WHY ATE?
via path analysis ?? Baron Kenny mess
week 2,3 even in RCT

moderation CATE (cnvrl)
individual diffs in response to intervention, heterogeneous treat
via pick-a-point (subgroups)
week 5 J-N region of significance
[ATE says all X in region, or no X]

---

even more . . .

> > My name is Patrick Forscher, and I am the lab instructor for a data
> analysis course in the UW-Madison Psychology Department. The instructor of
> record, Markus Brauer, and I are teaching the course in R. We have used
> your the mediate() function in the mediation package to demonstrate how to
> test for simple mediation (a la Baron & Kenny, 1986). However, we'd also
> like to teach the students to test for moderated mediation (when a
> mediation effect varies across levels of a third variable; Preacher,
> Rucker, & Hayes, 2007) and mediated moderation (when a variable provides
> the causal mechanism through which an interaction exerts its effect on a
> dependent variable; Muller, Judd, & Yzerbyt, 2005). Is the mediate()
> function able to test for moderated mediation or mediated moderation? If
> not, would you be able to recommend a package that can test for these cases?
> >
> > Thanks for your time and advice!

indiv diffs in mediation [M/F in framing anxiety]

why indiv diffs in effect? [aspirin] M/F

[Mediation-information] Using mediate() to test moderated mediation and mediated moderation

dustin tingley dtingley at gov.harvard.edu
Wed Nov 9 03:37:28 CET 2011

- Previous message: [Mediation-information] Using mediate() to test moderated mediation and mediated moderation
- Next message: [Mediation-information] question on mediation sensitivity analysis
- **Messages sorted by:** [date] [thread] [subject] [author]

Hi Patrick-
We currently do not include this functionality (interaction of mediation effect with a pre-treatment covariate). It is at the top of our list, thought. While not ideal, one can subset the data by the pre-treatment variable that you think is moderating the mediation. But that is a short term solution.

We'd love to hear examples/obtain data examples from you on this (beyond those papers which we are aware of). Can you please let us know where you go with this? When we adapt the program, it would be nice to benchmark this against what others provide.

thanks,
Dustin

Dustin Tingley
Government Department
Harvard University
http://scholar.harvard.edu/dtingley

On Tue, Nov 8, 2011 at 9:29 PM, Kosuke Imai <kimai at princeton.edu> wrote:

> *Dear Patrick,*
> *Thanks for your email. I think mediate() will accommodate the treatment-mediator interaction but the current version may not handle the moderated mediation effects. I'm ccing my coauthors to make sure that my answer is correct. We hope to include this functionality in the near future though. If interested, please subscribe to our mailing list where we announce the updates of the software and other relevant information.*
> *https://r-forge.r-project.org/projects/mediation/*
> *Best,*
> *Kosuke*
> *Department of Politics*
> *Princeton University*
> *http://imai.princeton.edu*
> *On Nov 8, 2011, at 1:16 PM, Patrick S Forscher wrote:*
> *> Hi Dr. Imai,*
> *> My name is Patrick Forscher, and I am the lab instructor for a data analysis course in the UW-Madison Psychology Department. The instructor of record, Markus Brauer, and I are teaching the course in R. We have used your the mediate() function in the mediation package to demonstrate how to test for simple mediation (a la Baron & Kenny, 1986). However, we'd also like to teach the students to test for moderated mediation (when a mediation effect varies across levels of a third variable; Preacher, Rucker, & Hayes, 2007) and mediated moderation (when a variable provides the causal mechanism through which an interaction exerts its effect on a dependent variable; Muller, Judd, & Yzerbyt, 2005). Is the mediate() function able to test for moderated mediation or mediated moderation? If not, would you be able to recommend a package that can test for these cases?*
> *> Thanks for your time and advice!*
> *> --*
> *> Patrick S Forscher*
> *> University of Wisconsin-Madison*

---------- next part ----------
An HTML attachment was scrubbed...
URL: <http://lists.r-forge.r-project.org/pipermail/mediation-information/attachments/20111108/91cd71ab/attachment.htm>

- Previous message: [Mediation-information] Using mediate() to test moderated mediation and mediated moderation
- Next message: [Mediation-information] question on mediation sensitivity analysis

PLOS ONE

RESEARCH ARTICLE

Predicting inadequate postoperative pain management in depressed patients: A machine learning approach

Arjun Parthipan[1], Imon Banerjee[2], Keith Humphreys[3,4], Steven M. Asch[4,5], Catherine Curtin[6,7], Ian Carroll[8☯], Tina Hernandez-Boussard[2,5,7☯*]

1 Department of Management Science and Engineering, Stanford University, Stanford, California, United States of America, **2** Department of Biomedical Data Sciences, Stanford University, Stanford, California, United States of America, **3** Department of Psychiatry, Stanford University, Stanford, California, United States of America, **4** VA Palo Alto Center for Innovation to Implementation, Palo Alto, California, United States of America, **5** Department of Medicine, Stanford University, Stanford, California, United States of America, **6** Department of Surgery, VA Palo Alto Healthcare System, Palo Alto, California, United States of America, **7** Department of Surgery, Stanford University, Stanford, California, United States of America, **8** Department of Anesthesiology, Stanford University, Stanford, California, United States of America

☯ These authors contributed equally to this work.
* boussard@stanford.edu

Citation: Parthipan A, Banerjee I, Humphreys K, Asch SM, Curtin C, Carroll I, et al. (2019) Predicting inadequate postoperative pain management in depressed patients: A machine learning approach. PLoS ONE 14(2): e0210575. https://doi.org/10.1371/journal.pone.0210575

Editor: Jose Palma, University of Murcia, SPAIN

Received: June 29, 2018

Accepted: December 26, 2018

Published: February 6, 2019

Data Availability Statement: The raw data used in this study contain personal health information and therefore can not be released to the public. However, all the code used to generate the analysis are available via GitHub (https://doi.org/10.25936/kjx8-0w74), including the details for the clinical phenotypes and the feature vectors, which are the de-identified data needed to run the algorithms.

Funding: This project was supported by grant number R01HS024096 from the Agency for Healthcare Research and Quality to TH. Dr. Humphreys was supported by a Senior Research

Abstract

Widely-prescribed prodrug opioids (e.g., hydrocodone) require conversion by liver enzyme CYP-2D6 to exert their analgesic effects. The most commonly prescribed antidepressant, selective serotonin reuptake inhibitors (SSRIs), inhibits CYP-2D6 activity and therefore may reduce the effectiveness of prodrug opioids. We used a machine learning approach to identify patients prescribed a combination of SSRIs and prodrug opioids postoperatively and to examine the effect of this combination on postoperative pain control. Using EHR data from an academic medical center, we identified patients receiving surgery over a 9-year period. We developed and validated natural language processing (NLP) algorithms to extract depression-related information (diagnosis, SSRI use, symptoms) from structured and unstructured data elements. The primary outcome was the difference between preoperative pain score and postoperative pain at discharge, 3-week and 8-week time points. We developed computational models to predict the increase or decrease in the postoperative pain across the 3 time points by using the patient's EHR data (e.g. medications, vitals, demographics) captured before surgery. We evaluate the generalizability of the model using 10-fold cross-validation method where the holdout test method is repeated 10 times and mean area-under-the-curve (AUC) is considered as evaluation metrics for the prediction performance. We identified 4,306 surgical patients with symptoms of depression. A total of 14.1% were prescribed both an SSRI and a prodrug opioid, 29.4% were prescribed an SSRI and a non-prodrug opioid, 18.6% were prescribed a prodrug opioid but were not on SSRIs, and 37.5% were prescribed a non-prodrug opioid but were not on SSRIs. Our NLP algorithm identified depression with a F1 score of 0.95 against manual annotation of 300 randomly sampled clinical notes. On average, patients receiving prodrug opioids had lower average pain scores ($p<0.05$), with the exception of the SSRI+ group at 3-weeks postoperative follow-up. However, SSRI+/Prodrug+ had significantly worse pain control at discharge, 3 and

Career Scientist Award from the Veterans Affairs Health Services Research and Development Service. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the Veterans Health Administration.

Competing interests: The authors have declared that no competing interests exist.

8-week follow-up ($p < .01$) compared to SSRI+/Prodrug- patients, whereas there was no difference in pain control among the SSRI- patients by prodrug opioid ($p>0.05$). The machine learning algorithm accurately predicted the increase or decrease of the discharge, 3-week and 8-week follow-up pain scores when compared to the pre-operative pain score using 10-fold cross validation (mean area under the receiver operating characteristic curve 0.87, 0.81, and 0.69, respectively). Preoperative pain, surgery type, and opioid tolerance were the strongest predictors of postoperative pain control. We provide the first direct clinical evidence that the known ability of SSRIs to inhibit prodrug opioid effectiveness is associated with worse pain control among depressed patients. Current prescribing patterns indicate that prescribers may not account for this interaction when choosing an opioid. The study results imply that prescribers might instead choose direct acting opioids (e.g. oxycodone or morphine) in depressed patients on SSRIs.

Introduction

Opioids are currently a first-line treatment of postoperative pain and surgery may be a gateway to opioid misuse [1,2]. Most surgical patients receive opioids, regardless of co-morbidities, prior opioid-related problems, or possible drug-drug interactions [3]. Depression is a common comorbidity and effects postoperative pain management. Longitudinal epidemiologic studies evaluating depression indicate that patients with depression are between two to five times more likely to have a new chronic pain problem at follow-up from one to eight years later [4–6]. Mental illness also increases the risk for opioid prescription abuse [7]. Antidepressants are the most commonly prescribed class of medications in the US and selective serotonin reuptake inhibitors (SSRIs) are the most commonly prescribed type of antidepressant. New evidence suggests that SSRI's could inhibit the metabolic conversion of certain opioids known as prodrug opioids, e.g. hydrocodone and codeine decreasing their efficacy for pain management. Therefore understanding the effect of SSRI antidepressants on postoperative pain management is essential as the US strives to move to a precision health system, and as opioid misuse and addiction continues to rise in the US during the current epidemic [8].

Opioids are thought to exert their analgesic effects by binding to the Mu opioid receptor in the brain and spinal cord. Some opioids directly bind to the mu-opioid receptor in their native form including oxycodone, morphine, hydromorphone, fentanyl, and methadone [9]. Other opioids require chemical conversion to an active form by a de-methylation reaction mediated in the human liver by CYP-2D6, a member of the cytochrome p450 enzyme system [10,11]. Such drugs are known as prodrug opioids, requiring metabolism and chemical modification to exert their pharmacological effect. Examples include hydrocodone- the most commonly prescribed drug in the nation which is the opioid ingredient in Vicodin, Lortab, and Norco. A recent study suggested that the interaction of a CYP-2D6 inhibitor might be important in reducing the effectiveness of hydrocodone [12]. Understanding the effects on pain control of the common antidepressant SSRI, a CYP-2D6 inhibitor, is essential to manage pain control is this vulnerable population with mental illness.

Our study's objective was to quantify the effect of the combined administration of prodrug opioids and SSRI antidepressant medication on postoperative pain among depressed patients undergoing surgical procedures. We hypothesized that patients taking SSRI who are prescribed a prodrug opioid will have worse postoperative pain. We further hypothesized that we

Common opioids less effective for patients on SSRI antidepressants

Patients taking antidepressants known as selective serotonin reuptake inhibitors do not respond well to hydrocodone, such as Vicodin, Stanford researchers report.

FEB 6
2019

Patients taking the most common form of antidepressant who are given the most widely prescribed opioid experience less pain relief, Stanford University School of Medicine investigators have discovered.

The finding could help combat the opioid epidemic, as poorly managed pain may lead to opioid abuse.

As many as 1 in 6 Americans takes antidepressants, mostly selective serotonin reuptake inhibitors.

Tina Hernandez-Boussard wants to help identify how to reduce exposure to opioids while still managing patients' pain.
Steve Fisch

Stanford researchers found that SSRIs reduce the effectiveness of hydrocodone and codeine, which are often prescribed to patients who have recently undergone surgery.

“This research is part of our effort to find ways to combat the opioid epidemic,” said Tina Hernandez-Boussard, PhD, MPH, associate professor of medicine, of biomedical data science and of surgery at Stanford. “We’re very interested in identifying how we can reduce opioid exposure while still managing patients’ pain.”

A paper describing the research was published Feb. 6 in *PLOS ONE*. Hernandez-Boussard and Ian Carroll, MD, assistant professor of anesthesiology, perioperative and pain medicine at Stanford, share senior authorship. The lead author is graduate student Arjun Parthipan.

Antidepressant inhibits enzyme

The researchers focused on the interaction between opioids and SSRIs because they knew that certain opioids, called prodrug opioids, need a liver enzyme to convert them into an active form that eases pain.

Ian Carroll

SSRIs inhibit this enzyme, so the researchers hypothesized that patients taking SSRIs in combination with prodrug opioids would receive less pain relief. Prodrug opioids include codeine and hydrocodone, which is sold under brand names such as Vicodin, Lorcet and Lortab. SSRIs include Prozac, Paxil, Zoloft and Celexa.

For the study, the research team analyzed de-identified data for 4,306 surgical patients with a diagnosis or symptoms of depression in the electronic health records at Stanford Health Care.

The researchers found that the patients on SSRIs who were prescribed prodrug opioids experienced more pain when they left the hospital, as well as three weeks later and eight weeks later. The patients on both SSRIs and prodrug opioids registered nearly one point more pain on a scale of 1 to 10 than the patients not on SSRIs who were prescribed prodrug opioids.

Algorithm predicts opioid response

The researchers built a machine-learning algorithm that predicts how a patient will respond to different types of opioids. The algorithm is available online.

The study focused on surgery patients because they nearly always receive opioid prescriptions, yet the findings apply to anyone taking short-term opioids.

Carroll noted that hydrocodone is the most frequently prescribed drug in the nation. With SSRIs the most frequently prescribed class of drugs, he said, the chance that any patient will be on both drugs is high.

There's no proof that better pain management reduces the number of opioid overdoses, cautioned Carroll. But poor pain control has been shown to be a risk factor for chronic pain, and it may lead to more prolonged opioid use, along with misuse.

"Presumably, every day you take opioids, the risk you'll abuse it increases," Carroll said.

The authors concluded that to manage pain for patients on SSRIs, prescribers should choose nonopioids or direct-acting opioids. Direct-acting opioids, which include morphine, fentanyl and oxycodone, do not need the liver enzyme to convert the drug into a form that eases pain.

Prescribers typically choose hydrocodone or codeine because of a perception that they are milder than hydromorphone, whose trade names include Dilaudid and Exalgo, or morphine, Carroll said.

"The prescribing of hydrocodone has more to do with history and perception," he said. "The liver converts hydrocodone into hydromorphone and converts codeine into morphine,so the result is the same."

Carroll added that depressed patients' complaints about pain after receiving opioids are often dismissed because of their mental state, when the problem lies in an unfortunate drug interaction.

> *"Depressed patients are at greater risk for pain, and we're failing them because we're not educated enough about the drugs we're giving them."*

"Depressed patients are at greater risk for pain, and we're failing them because we're not educated enough about the drugs we're giving them," he said.

Hernandez-Boussard is a member of Stanford Bio-X, and she and Carroll are members of the Stanford Cancer Institute.

Other Stanford co-authors of the study are research scientist Imon Banerjee, PhD; Keith Humphreys, PhD, professor of psychiatry and behavioral sciences and the Esther Ting Memorial Professor; Steven Asch, MD, MPH, professor of medicine; and Catherine Curtin, MD, associate professor of surgery.

The work was funded by the Agency for Health Care Research and Quality.

Stanford's departments of Medicine, of Biomedical Data Science, of Surgery and of Anesthesiology, Perioperative and Pain Medicine also supported the work.

By
MANDY ERICKSON
Mandy Erickson is a science writer for the medical school's Office of Communication & Public Affairs.

Stanford Medicine integrates research, medical education and health care at its three institutions - Stanford University School of Medicine, Stanford Health Care (formerly Stanford Hospital & Clinics), and Lucile Packard Children's Hospital Stanford. For more information, please visit the Office of Communication & Public Affairs site at http://mednews.stanford.edu.

effects.[23] As a result, in the intention-to-treat analysis that examined the overall effect of treatment, the effect of medical management may have been understated.

The **Table** summarizes these definitions, but also fills in other possible relationships among target measure, treatment choice, and outcome. For example, if a posttreatment variable (not a moderator of treatment) is not correlated with treatment choice (also not a mediator of treatment) but has an interactive effect with treatment on the outcome, then treatment is a moderator of that measure (not vice versa). Thus, for example, in a program designed to treat depression, the effect of the death of a relative or friend during the treatment (presumably not related to which treatment was assigned) on outcome may be moderated by treatment. This would be the case if those in T are better able to cope with such an event than those in C. The directionality of mediation and moderation is important to note. Moderators always precede what they moderate, which in turn precedes outcome; mediators always come between what they mediate and the outcome.

A baseline measure (not a mediator) that has a main effect on outcome but no interactive effect (not a moderator) might be called a nonspecific predictor of outcome. Such a target measure predicts response in both treatment groups, but the effect size of treatment is the same regardless of the value of the target measure. Thus, for example, in the multisite ADHD study, there were major site differences in almost all outcomes, but few site-by-treatment interactive effects.[24] Similar results were obtained in a multisite RCT that tested the relative effectiveness of CBT and interpersonal psychotherapy (IPT) for bulimia nervosa.[25] In both cases, site was a nonspecific predictor of outcome, but not a moderator (ie, the effect size of treatment did not differ over sites, even though the response to treatment did).

A posttreatment measure (not a moderator) uncorrelated with treatment (not a mediator) that has a main effect but no interaction might also be called a nonspecific predictor of outcome. Unlike the ADHD result, compliance with treatment might nondifferentially enhance treatment response to all treatments. If so, compliance with treatment would be a nonspecific predictor of outcome. The message would be that improvement of compliance would enhance outcomes whatever the treatment.

A posttreatment measure (not a moderator) that is correlated with treatment, which has neither a main nor an interactive effect with treatment on outcome, is an independent outcome of treatment. For example, a cardiovascular risk reduction program may reduce weight and increase activity level, but it may be that the decrease in weight does not relate to the increase in activity level or vice versa. In such a case, weight decrease and activity increase are 2 independent outcomes of the treatment.

Finally, a target variable, either before or after baseline uncorrelated with treatment, that has neither a main nor an interactive effect on an outcome is not demonstrated to be relevant to the treatment outcome. In a sample from the population, this may, of course, be a matter of inadequate power owing to small sample size or unreliable measurement.

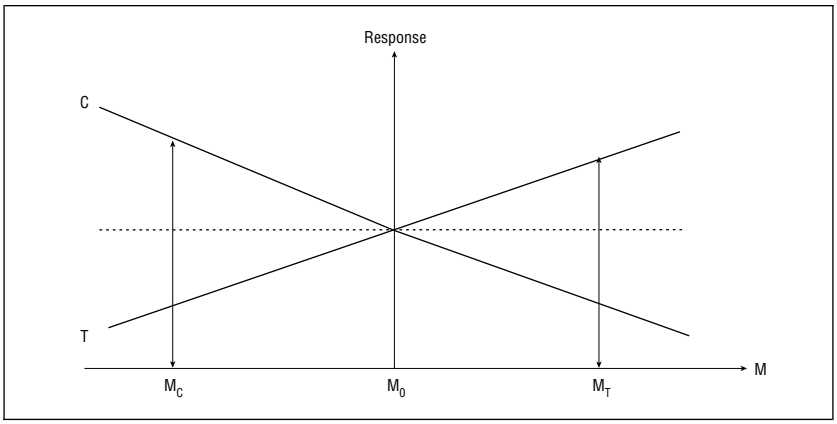


Figure 2. A special case in which there is no main effect of treatment, no main effect of moderator or mediator (M), and no overall effect of treatment, but in which treatment may change not only the level but also the action of M on the outcome, a mediating effect. T indicates treatment group; C, control or comparison group; M_C, the mean of M in C; M_T, the mean of M in T; and M_0, the midpoint of these two.

COMMENT

The Role of Theory

Clearly, there should be some theoretical basis for the choice of measures to be considered as possible mediators and moderators. For example, a measure that is simply part of the definition of one of the treatments should not be considered as either a mediator or a moderator. When one considers comparing the effects of a medication treatment vs a psychotherapy treatment, one

Summary of Population Definitions Relating Target Measure to Treatment and Outcome

Target Measure	Correlation With Treatment	Relationship to Outcome in Linear Model	Classification of Target Measure
Pretreatment	No (by definition)	Interaction with or without main effect	Moderator of treatment outcome
Pretreatment	No (by definition)	Main effect only	Nonspecific predictor of treatment outcome
Posttreatment	Yes	Main effect or interaction	Mediator of treatment outcome
Posttreatment	Yes	Neither main effect nor interaction	Independent outcome of treatment
Posttreatment	No	Interaction with or without main effect	Treatment moderates target variable
Posttreatment	No	Main effect only	Nonspecific predictor of treatment outcome
Pretreatment or posttreatment	No	Neither main effect nor interaction	Target measure irrelevant to treatment outcome

R Multilevel package

sobel *Estimate Sobel's (1982) Test for Mediation*

Description

Estimate Sobel's (1982) indirect test for mediation. The function provides an estimate of the magnitude of the indirect effect, Sobel's first-order estimate of the standard error associated with the indirect effect, and the corresponding z-value. The estimates are based upon three models as detailed on page 84 of MacKinnon, Lockwood, Hoffman, West and Sheets (2002).

Usage

```
sobel(pred,med,out)
```

Arguments

pred	The predictor or independent variable.
med	The mediating variable.
out	The outcome or dependent variable.

Value

Model.1	A summary of coefficients from Model 1 of MacKinnon et al., (2002).
Model.2	A summary of coefficients from Model 2 of MacKinnon et al., (2002).
Model.3	A summary of coefficients from Model 3 of MacKinnon et al., (2002).
Indirect.Effect	The estimate of the indirect mediating effect.
SE	Sobel's (1982) Standard Error estimate.
z.value	The estimated z-value.
N	The number of observations used in model estimation.

Author(s)

Paul Bliese ⟨paul.bliese@us.army.mil⟩

References

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7, 83-104.

Sobel, M. E., (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), Sociological Methodology 1982 (pp. 290-312). Washington, DC: American Sociological Association.

See Also

sobel.lme

Examples

```
data(bh1996)

#A small but significant indirect effect indicates leadership mediates
#the relationship between work hours and well-being.
sobel(pred=bh1996$HRS,med=bh1996$LEAD,out=bh1996$WBEING)
```

---

sobel.lme *Estimate Sobel's (1982) Test for Mediation in Two-Level lme Model*

Description

Estimate Sobel's (1982) indirect test for mediation in a two-level linear mixed effects (lme) model. The function provides an estimate of the magnitude of the indirect effect, Sobel's first-order estimate of the standard error associated with the indirect effect, and the corresponding z-value. The estimates are based upon three models as detailed on page 84 of MacKinnon, Lockwood, Hoffman, West and Sheets (2002). By estimating the three models in a two-level lme framework, one can account for the effects of hierarchical nesting in the data, and thereby obtain more accurate estimates of standard errors (see Bliese & Hanges, 2004).

Usage

```
sobel.lme(pred, med, out, grpid)
```

Arguments

pred	The predictor or independent variable.
med	The mediating variable.
out	The outcome or dependent variable.
grpid	A variable identifying the groups in which pred, med and out are nested.

Value

Model.1	Summary of coefficients from Model 1 of MacKinnon et al.
Model.2	Summary of coefficients from Model 2 of MacKinnon et al.
Model.3	Summary of coefficients from Model 3 of MacKinnon et al.
Indirect.Effect	Estimate of indirect mediating effect.
SE	Sobel's (1982) Standard Error estimate.
z.value	The estimated z-value.
N	The number of observations used in model estimation.
n.grps	The number of groups used in model estimation.

Author(s)

Paul Bliese ⟨paul.bliese@us.army.mil⟩

References

Bliese, P. D. & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. Organizational Research Methods, 7, 400-417.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7, 83-104.

Sobel, M. E., (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), Sociological Methodology 1982 (pp. 290-312). Washington, DC: American Sociological Association.

See Also

sobel

Examples

```
data(bh1996)
library(nlme)

#A small but significant indirect effect indicates leadership mediates
#the relationship between work hours and well-being.

with(bh1996, sobel.lme(pred=HRS,med=LEAD,out=WBEING,grpid=GRP))
```

---

summary.agree.sim *S3 method for class 'agree.sim'*

Description

This function provides a concise summary of objects created using the functions rwg.sim and rwg.j.sim.

Usage

```
## S3 method for class 'agree.sim':
summary(object,...)
```

Arguments

object	An object of class 'agree.sim'.
...	Optional additional arguments. None used.

Package 'MBESS'

February 15, 2013

Title MBESS

Version 3.3.3

Date 2012-12-15

Author Ken Kelley and Keke Lai

Maintainer Keke Lai <LaiKeke@ASU.edu>

Depends R (>= 2.6.0)

Suggests MASS, sem, boot, nlme, gsl, lavaan, parallel, snow, OpenMx

Description MBESS implements methods that are especially useful to researchers working within the behavioral, educational, and social sciences (both substantive researchers and methodologists), Many of the methods contained within MBESS are applicable to quantitative research in general,

License GPL (>= 2)

URL http://nd.edu/~kkelley/site/MBESS.html

Repository CRAN

Date/Publication 2012-12-17 06:09:14

NeedsCompilation no

R topics documented:

Details

Package:	MBESS
Type:	Package
Version:	3.3.3
Date:	2012-12-15
License:	GPL(>=2)

Please read the manual and visit the corresponding web site http://nd.edu/~kkelley/site/MBESS.html for information on the capabilities of the MBESS package. Feel free to contact me if there is a feature you would like to see added if it would complement the goals of the MBESS package.

Author(s)

Ken Kelley <<KKelley@ND.Edu>; http://www.nd.edu/~kkelley> and Keke Lai <LaiKeke@ASU.edu>

Maintainer: Keke Lai <LaiKeke@ASU.edu>; Ken Kelley <<KKelley@ND.Edu>; http://www.nd.edu/~kkelley>

`mediation` *Effect sizes and confidence intervals in a mediation model*

Description

Automate the process of simple mediation analysis (one independent variable and one mediator) and effect size estimation for mediation models, as discussed in Preacher and Kelley (2010).

Usage

```
mediation(x, mediator, dv, S = NULL, N = NULL, x.location.S = NULL,
mediator.location.S = NULL, dv.location.S = NULL, mean.x = NULL,
mean.m = NULL, mean.dv = NULL, conf.level = 0.95,
bootstrap = FALSE, B = 1000, which.boot="both", save.bs.replicates=FALSE)
```

Arguments

`x`	vector of the predictor/independent variable
`mediator`	vector of the mediator variable
`dv`	vector of the dependent/outcome variable
`S`	Covariance matrix
`N`	Sample size, necessary when a covariance matrix (`S`) is used
`x.location.S`	location of the predictor/independent variable in the covariance matrix (`S`)
`mediator.location.S`	location of the mediator variable in the covariance matrix (`S`)

`dv.location.S` location of the dependent/outcome variable in the covariance matrix (`S`)

`mean.x` mean of the `x` (independent/predictor) variable when a covariance matrix (`S`) is used

`mean.m` mean of the `m` (mediator) variable when a covariance matrix (`S`) is used

`mean.dv` mean of the `y/dv` (dependent/outcome) variable when a covariance matrix (`S`) is used

`conf.level` desired level of confidence (e.g., .90, .95, .99, etc.)

`bootstrap` TRUE or FALSE, based on whether or not a bootstrap procedure is performed to obtain confidence intervals for the various effect sizes

`B` number of bootstrap replications when `bootstrap=TRUE` (e.g., 10000)

`which.boot` which bootstrap method to use. It can be `Percentile` or `BCa`, or both

`save.bs.replicates`
Logical argument indicating whether to save the each bootstrap sample or not

Details

Based on the work of Preacher and Kelley (2010), this function implements (simple) mediation analysis in a way that automates much of the results that are generally of interest. More specifically, three regression outputs are automated as is the calculation of effect sizes that are thought to be useful or potentially useful in the context of mediation. Much work on mediation models exists in the literature, which should be consulted for proper interpretation of the effect sizes, models, and meaning of results.

Value

`Y.on.X$Regression.Table`
Regression table of `Y` conditional on `X`

`Y.on.X$Model.Fit`
Summary of model fit for the regression of `Y` conditional on `X`

`M.on.X$Regression.Table`
Regression table of `X` conditional on `M`

`M.on.X$Model.Fit`
Summary of model fit for the regression of `X` conditional on `M`

`Y.on.X.and.M$Regression.Table`
Regression table of `Y` conditional on `X` and `M`

`Y.on.X.and.M$Model.Fit`
Summary of model fit for the regression of `Y` conditional on `X` and `M`

`Indirect.Effect`
the product of $\hat{a} \times \hat{b}$, where $\hat{a}$ and $\hat{b}$ are the estimated coefficients of the path from the independent variable to the mediator and the path from the mediator to the dependent variable

`Indirect.Effect.Partially.Standardized`
It is the indirect effect (see `Indirect.Effect` above) divided by the estimated standard deviation of `Y` (MacKinnon, 2008)

Package 'powerMediation'

April 29, 2013

Version 0.1.3

Date 2013-04-29

Title Power/Sample size calculation for mediation analysis

Author Weiliang Qiu <stwxq@channing.harvard.edu>

Maintainer Weiliang Qiu <stwxq@channing.harvard.edu>

Depends R (>= 2.9.0), stats

Description The package contains functions (1) for calculating power,sample size, and minimal detectable mediation effect for testing mediation effect in linear, logistic, Poisson, or cox regression; (2) for calculating power, sample size, and minimal detectable slope for testing the slope in a simple linear regression (only one predictor); (3) for calculating power, and sample size for testing odds ratio in a simple logistic regression with continuous predictor or binary predictor.

License GPL (>= 2)

Repository CRAN

Date/Publication 2013-04-29 16:56:41

NeedsCompilation no

R topics documented: