Evidence-based Decision Making
Alternatives to Experiments

Rui Mata, FS 2023

Version: April 24th, 2023

Die Fakultät für Psychologie der Universität Basel lädt Sie ein!

DIENSTAG, 9. MAI 2023, 17:00

INFORMATIONSVERANSTALTUNG

ZUM MASTERSTUDIUM IN SOZIAL-, WIRTSCHAFTS- UND ENTSCHEIDUNGSPSYCHOLOGIE

17:00 Uhr
FAKULTÄT FÜR PSYCHOLOGIE
MISSIONSSTRASSE 62A
HÖRSAAL 00.006

DIENSTAG, 9. MAI 2023, 17:30

PSYCHOLOGIE IN DER PRAXIS

ABSOLVENTEN/INNEN DER MASTERVERTIEFUNGSRICHTUNG SOZIAL-, WIRTSCHAFTS- UND ENTSCHEIDUNGSPSYCHOLOGIE BERICHTEN VON IHREN BERUFSERFAHRUNGEN NACH DEM STUDIUM

MIT ANSCHLIESSENDEM APÉRO

17:30 Uhr
FAKULTÄT FÜR PSYCHOLOGIE
MISSIONSSTRASSE 62A
HÖRSAAL 00.006

Goals

- Understand the nature of causal inference as the comparison of treatment to some counterfactual
- List different methods of causal inference (e.g., randomization/experiments, regression, regression discontinuity) and associated limitations

Does education work?

Quasi-Experimental Designs: Educational effects on intelligence

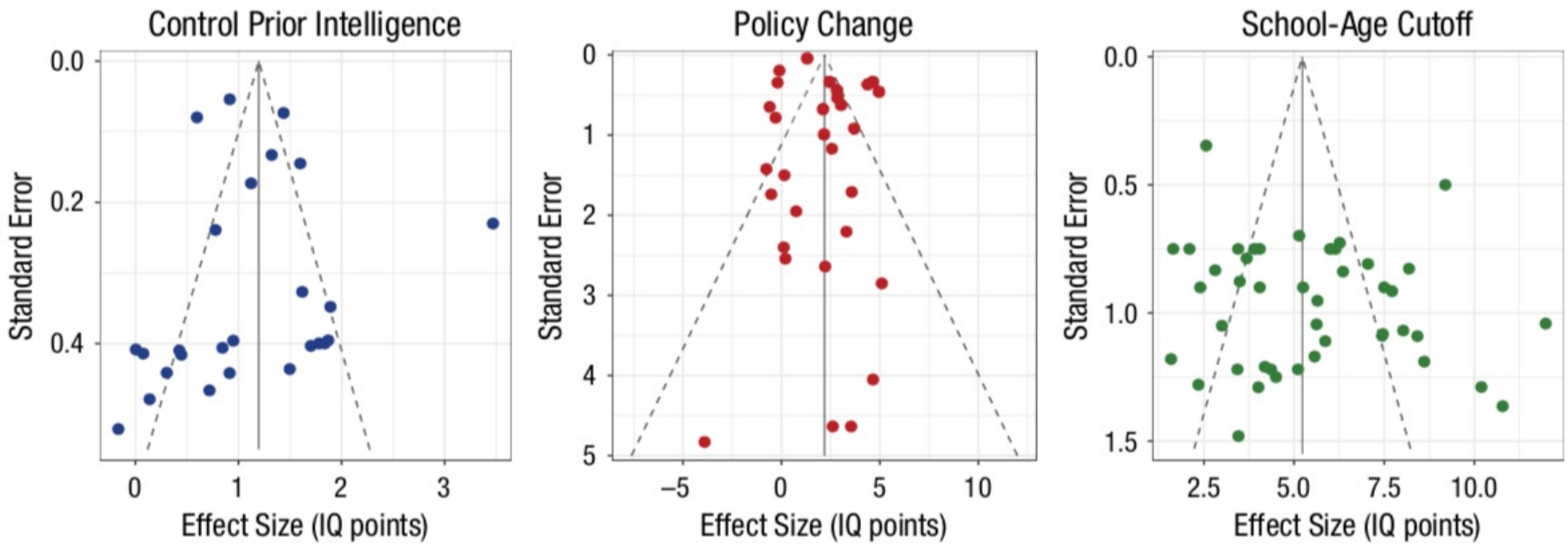


Fig. 2. Funnel plots showing standard error as a function of effect size, separately for each of the three study designs. The dotted lines form a triangular region (with a central vertical line showing the mean effect size) where 95% of estimates should lie in the case of zero within-group heterogeneity in population effect sizes. Note that 42 of the total 86 standard errors reported as approximate or as averages in the original studies were not included for the school-age-cutoff design.

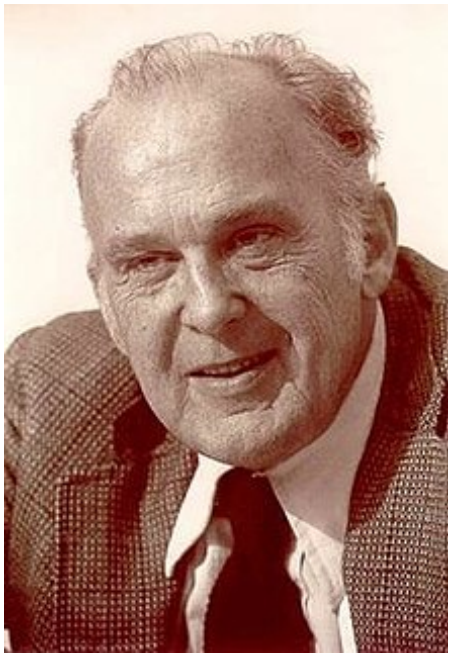
control prior intelligence = longitudinal studies in which cognitive testing data were collected before and after variation in the duration of education (e.g., before and after university vs. no university)

policy change = study of the effects of a change in educational duration (e.g., increase of compulsory education by 1 year) on mental testing

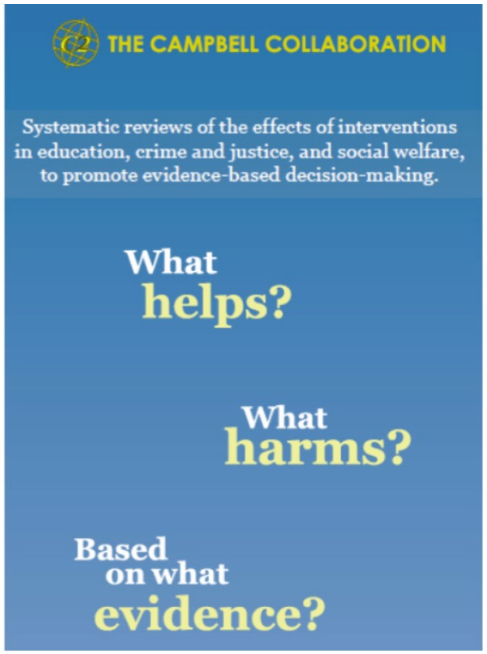
school-age cutoff = studies use regression-discontinuity analysis to leverage the fact that school districts implement a date-of-birth cutoff for school entry (example: compare 3.9-year olds that did not attend "Kindsgi" vs. 4.0 year-olds that did)

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*(8), 1358–1369. http://doi.org/10.1177/0956797618774253

There are alternatives…



Donald Campbell
1916-1996



THE CAMPBELL COLLABORATION

Quasi-experimental designs

Before-and-after measures

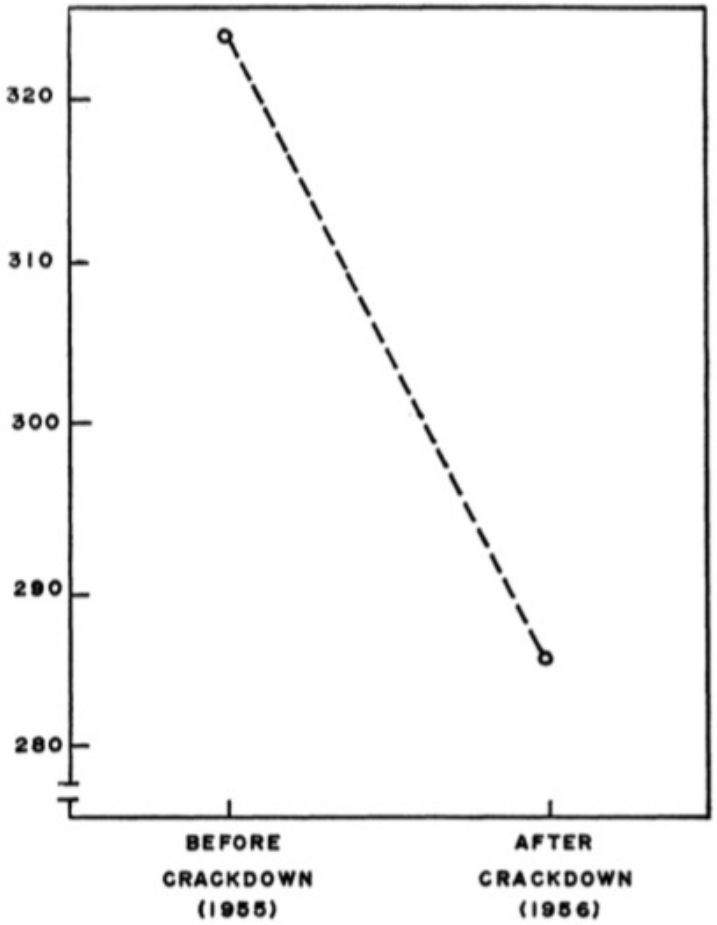


Figure 1. Connecticut Traffic Fatalities, 1955-1956

- was 1956 a dry year? (history)
- overall trends in road safety? (maturation)
- did publicising of death rates have an effect? (testing)
- were fatalities counted differently? (instrumentation)
- was this a big decrease? (instability)
- was 1995 an extreme year? (regression)

Campbell, D. T., Ross, H. L. (1968). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review, 3(1), 33. http://doi.org/10.2307/3052794

Quasi-experimental designs

Interrupted time series

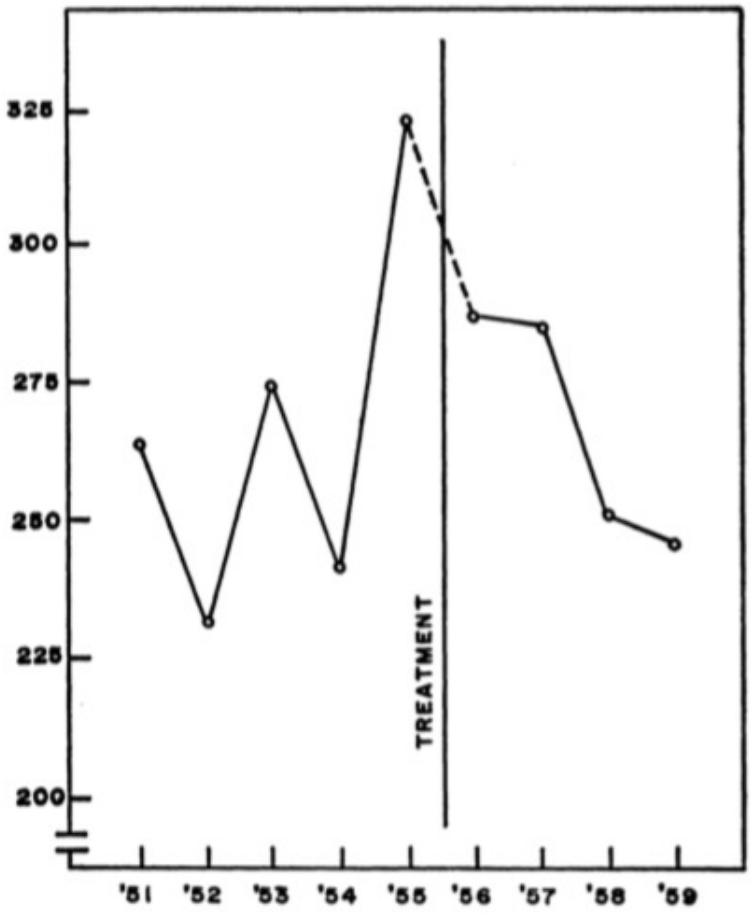


Figure 2. Connecticut Traffic Fatalities, 1951-1959

- was publicising of death rates similar across years? (testing)
- were fatalities counted differently before and after the intervention? (instrumentation)

Campbell, D. T., Ross, H. L. (1968). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review, 3(1), 33. http://doi.org/10.2307/3052794

Quasi-experimental designs

Multiple time series

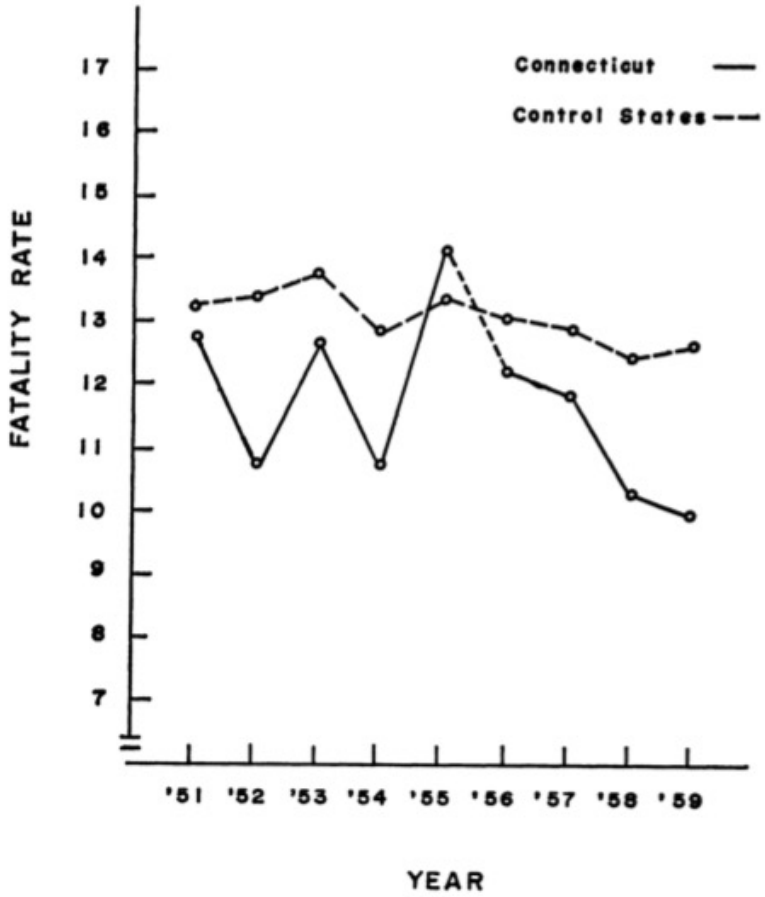


Figure 3. Connecticut and Control States Traffic Fatalities, 1951-1959 (per 100,000 population)

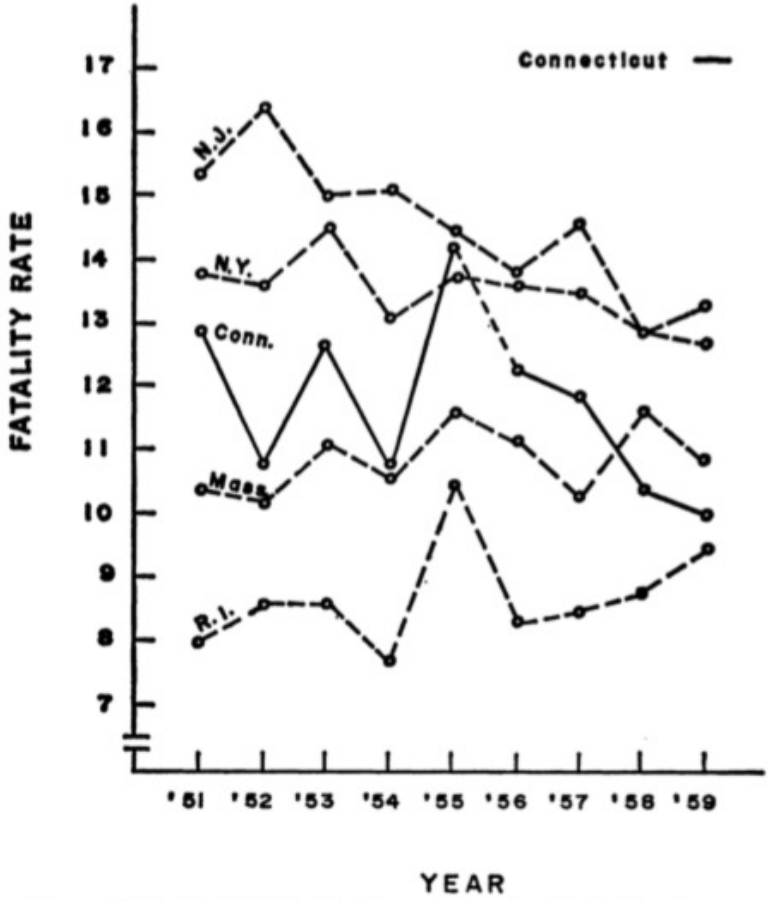
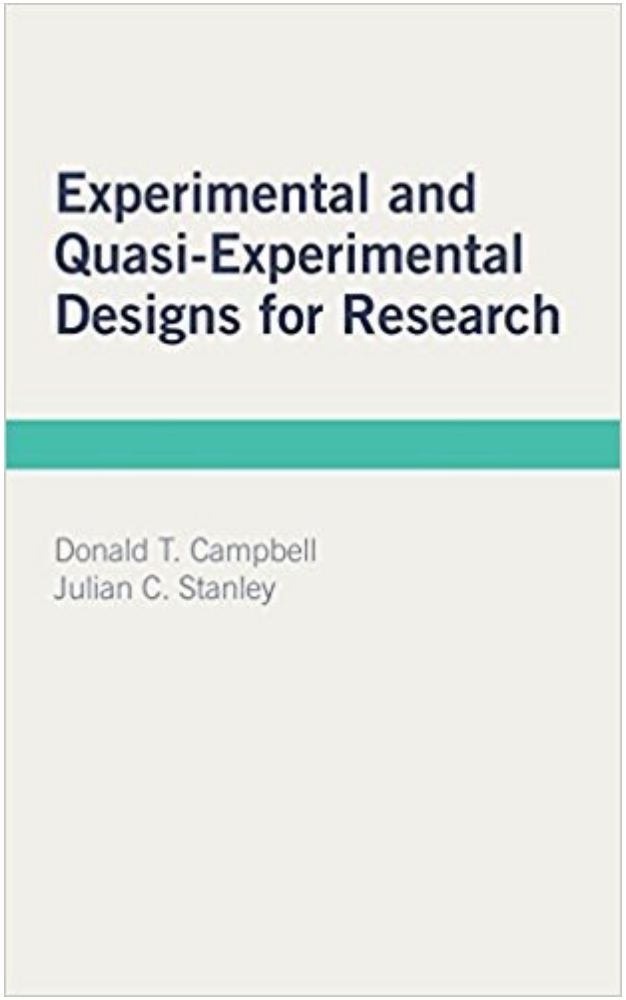


Figure 4. Traffic Fatalities for Connecticut, New York, New Jersey, Rhode Island, and Massachusetts (per 100,000 persons)

Campbell, D. T., Ross, H. L. (1968). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review, 3(1), 33. http://doi.org/10.2307/3052794


Experimental and
Quasi-Experimental
Designs for Research
Donald T. Campbell
Julian C. Stanley

1963

TABLE 1

SOURCES OF INVALIDITY FOR DESIGNS 1 THROUGH 6

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Pre-Experimental Designs:												
1. One-Shot Case Study X O	−	−				−	−			−		
2. One-Group Pretest-Posttest Design O X O	−	−	−	−	?	+	+	−		−	−	?
3. Static-Group Comparison X O --- O	+	?	+	+	+	−	−	−			−	
True Experimental Designs:												
4. Pretest-Posttest Control Group Design R O X O R O O	+	+	+	+	+	+	+	+		−	?	?
5. Solomon Four-Group Design R O X O R O O R X O R O	+	+	+	+	+	+	+	+		+	?	?
6. Posttest-Only Control Group Design R X O R O	+	+	+	+	+	+	+	+		+	?	?

Note: In the tables, a minus indicates a definite weakness, a plus indicates that the factor is controlled, a question mark indicates a possible source of concern, and a blank indicates that the factor is not relevant.

It is with extreme reluctance that these summary tables are presented because they are apt to be "too helpful," and to be depended upon in place of the more complex and qualified presentation in the text. No + or − indicator should be respected unless the reader comprehends why it is placed there. In particular, it is against the spirit of this presentation to create uncomprehended fears of, or confidence in, specific designs.

Campbell & Stanley (1963)

TABLE 2

SOURCES OF INVALIDITY FOR QUASI-EXPERIMENTAL DESIGNS 7 THROUGH 12

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Quasi-Experimental Designs:												
7. Time Series O O O OXO O O O	−	+	+	?	+	+	+	+	−	?	?	
8. Equivalent Time Samples Design X_1O X_2O X_1O X_2O, etc.	+	+	+	+	+	+	+	+	−	?	−	−
9. Equivalent Materials Samples Design M_aX_1O M_bX_2O M_cX_1O M_dX_2O, etc.	+	+	+	+	+	+	+	+	−	?	?	−
10. Nonequivalent Control Group Design O X O O O	+	+	+	+	?	+	+	−	−	?	?	
11. Counterbalanced Designs X_1O X_2O X_3O X_4O X_2O X_4O X_1O X_3O X_3O X_1O X_4O X_2O X_4O X_3O X_2O X_1O	+	+	+	+	+	+	+	?	?	?	?	−
12. Separate-Sample Pretest-Posttest Design R O (X) R X O	−	−	+	?	+	+	−	−	+	+	+	
12a. R O (X) R X O R O (X) R X O	+	−	+	?	+	+	−	+	+	+	+	
12b. R O_1 (X) R O_2 (X) R X O_3	−	+	+	?	+	+	−	?	+	+	+	
12c. R O_1 X O_2 R X O_3	−	−	+	?	+	+	+	−	+	+	+	

Campbell & Stanley (1963)

TABLE 3

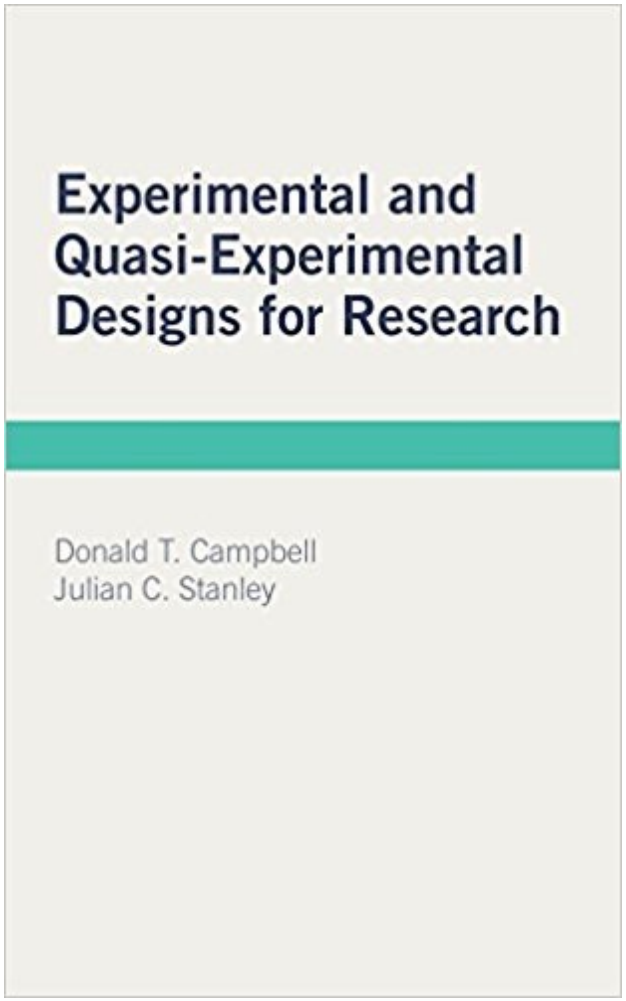
SOURCES OF INVALIDITY FOR QUASI-EXPERIMENTAL DESIGNS 13 THROUGH 16

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Quasi-Experimental Designs Continued:												
13. Separate-Sample Pretest-Posttest Control Group Design R O (X) R X O --- R O R O	+	+	+	+	+	+	+	−	+	+	+	
13a. R O (X) R X O --- R O (X) R X O --- R O (X) R X O R O R O --- R O R O --- R O R O	+	+	+	+	+	+	+	+	+	+	+	
14. Multiple Time-Series O O O X O O O --- O O O O O O O	+	+	+	+	+	+	+	+	−	−	?	
15. Institutional Cycle Design Class A X O_1 --- Class B_1 R O_2 X O_3 Class B_2 R X O_4 --- Class C O_5 X ===												
[a]Gen. Pop. Con. Cl. B O_6												
[a]Gen. Pop. Con. Cl. C O_7												
$O_2 < O_1$ $O_5 < O_4$	+	−	+	+	?	−	?		+	?	+	
$O_2 < O_3$	−	−	−	?	?	+	+		−	?	+	
$O_2 < O_4$	−	−	+	?	?	+	?		+	?	?	
$O_6 = O_7$ $O_{2y} = O_{2o}$		+						−				
16. Regression Discontinuity	+	+	+	?	+	+	?	+	+	−	+	+

[a] General Population Controls for Class B, etc.

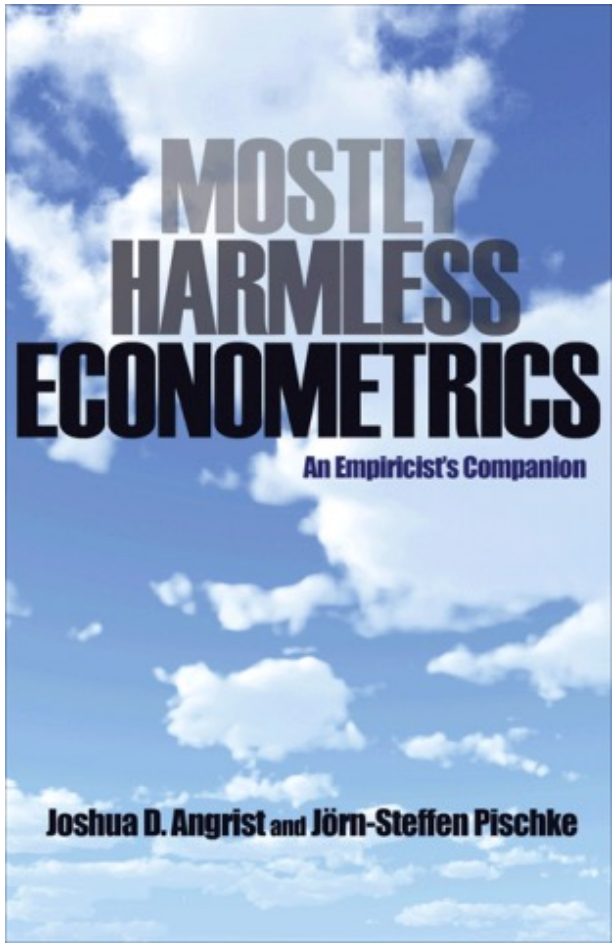
Campbell & Stanley (1963)

Experimental and Quasi-experimental Designs



"In conclusion, in this chapter we have discussed alternatives in the arrangement or design of experiments, with particular regard to the problems of control of extraneous variables and threats to validity. (…) Through out, attention has been called to the possibility of **creatively** utilizing the idiosyncratic features of any specific research situation in designing unique tests of causal hypotheses.

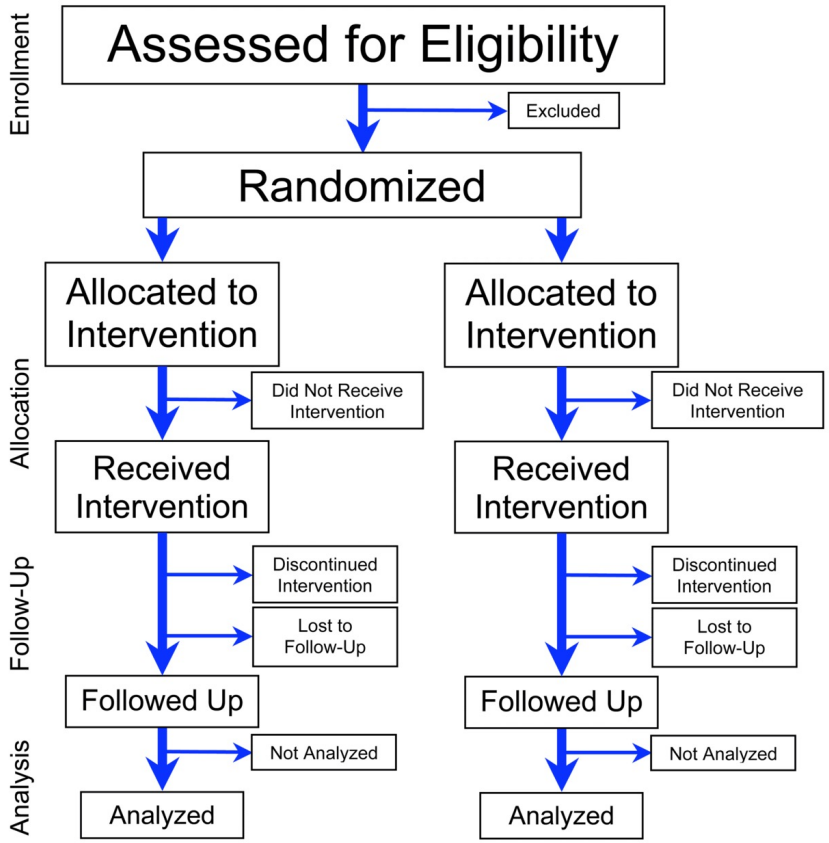
“Furious Five” statistical methods for causal inference



- Randomisation
- Regression
- Instrumental variables
- Difference in differences
- Regression discontinuity

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. http://doi.org/10.1257/jep.24.2.3

Randomisation



Full randomisation is seldom available in practice…

> The "ideal" data, from the viewpoint of the analyst, would be data from an incompetent advertiser who allocated expenditures randomly across cities. If ad expenditure is truly random, then we do not have to worry about confounding variables because the predictors will automatically be orthogonal to the error term. However, statisticians are seldom lucky enough to have a totally incompetent client.

Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

Regression

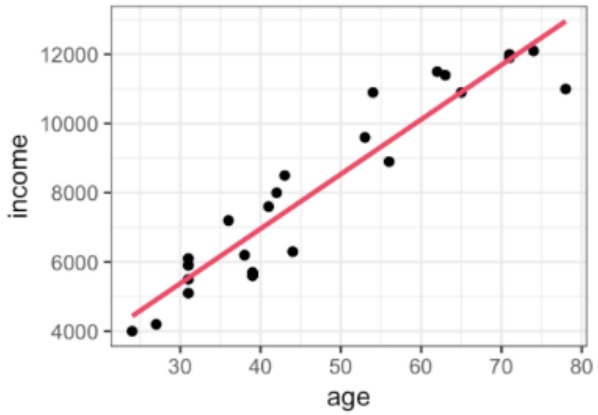
Regression analysis is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable (criterion) and one or more independent variables (predictors). More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are fixed.

Regression

Simple Linear Regression

Definition: Simple linear regression is a linear model with one predictor x, and where the error term ϵ is Normally distributed.

$$y = \beta_0 + \beta_1 x + \epsilon$$



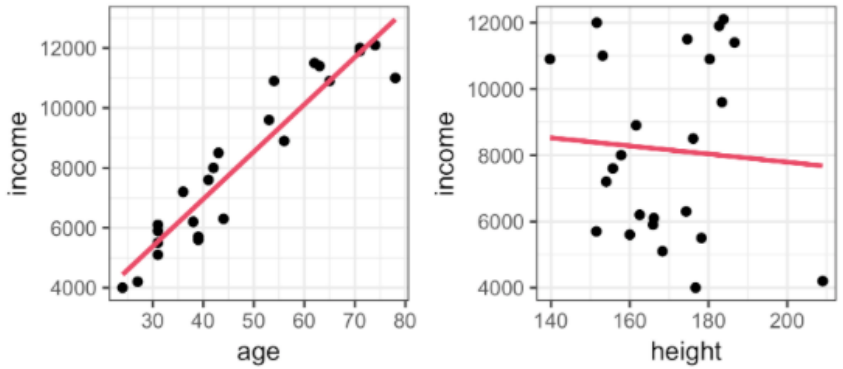
Regression

Multiple Linear Regression

Definition: Multiple linear regression is a linear model with many predictors $x_1, x_2, \ldots x_n$, and where the error term ϵ is Normally distributed.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

Parameter	Description	In words
β_0	Intercept	When all x values are 0, what is the predicted value for y?
$\beta_1, \beta_2, \ldots$	Coefficient for $x_1, x_2, \ldots$	For every increase of 1 in coefficient for $x_1, x_2, \ldots$ how does y change?



Formula

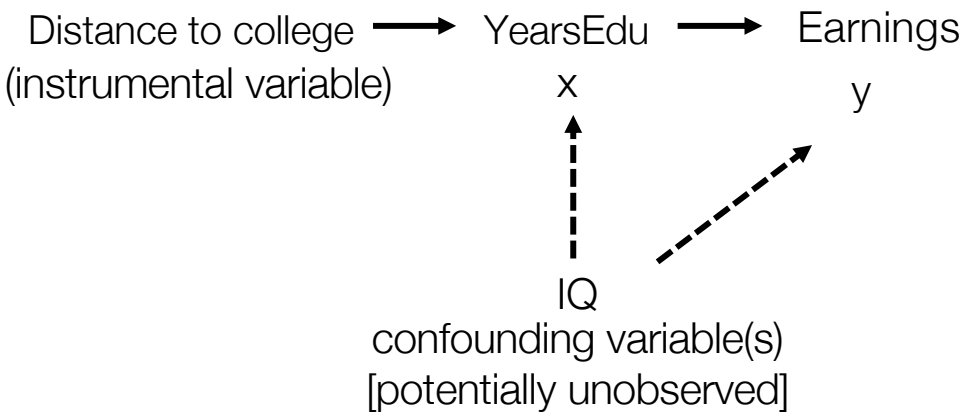
$$income = 1628 + 147 \times age - 4.1 \times height + \epsilon$$

Coefficients

$$\beta_0 = 1628, \beta_{age} = 147, \beta_{weight} = -4.1$$

Instrumental variables

The method of instrumental variables (IV) is used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomized experiment. Intuitively, the method is used when an explanatory variable of interest is correlated with the error term, in which case ordinary least squares gives biased results. A valid instrument (instrumental variable) induces changes in the explanatory variable (x) but has no independent effect on the dependent variable (y), allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.



Estimation though two-stage least squares.

Stage 1: generate predictions of YearsEdu:

$YearsEdu\_pred = B_0 + B_1\ DisttoCollege + Error$

Stage 2: test whether YearsEdu_pred is significantly associated with earnings:

$Earnings = B_0 + B_1\ YearsEdu\_pred + Error$

Problem: Good instrumental variables (i.e., that are correlated with x but not any confounding variables) are hard to find…

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. ***Journal of Economic Perspectives, 15***(4), 69–85.

Instrumental variables

Table 1

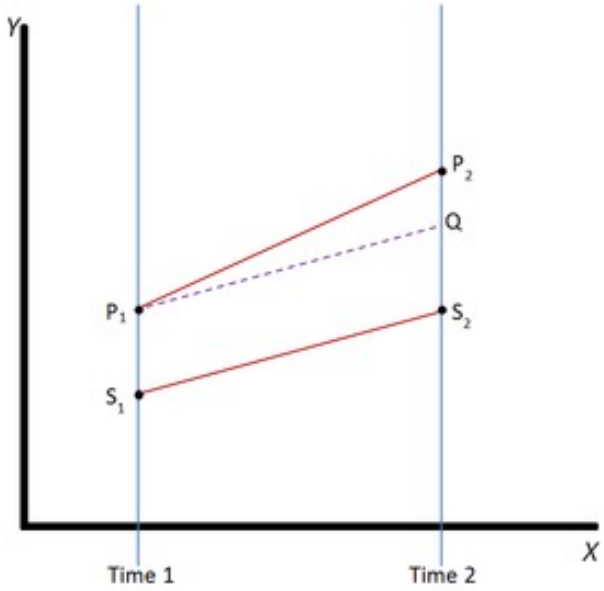
Examples of Studies That Use Instrumental Variables to Analyze Data From Natural and Randomized Experiments

Outcome Variable	*Endogenous Variable*	*Source of Instrumental Variable(s)*	*Reference*
		1. Natural Experiments	
Labor supply	Disability insurance replacement rates	Region and time variation in benefit rules	Gruber (2000)
Labor supply	Fertility	Sibling-Sex composition	Angrist and Evans (1998)
Education, Labor supply	Out-of-wedlock fertility	Occurrence of twin births	Bronars and Grogger (1994)
Wages	Unemployment insurance tax rate	State laws	Anderson and Meyer (2000)
Earnings	Years of schooling	Region and time variation in school construction	Duflo (2001)
Earnings	Years of schooling	Proximity to college	Card (1995)
Earnings	Years of schooling	Quarter of birth	Angrist and Krueger (1991)
Earnings	Veteran status	Cohort dummies	Imbens and van der Klaauw (1995)
Earnings	Veteran status	Draft lottery number	Angrist (1990)
Achievement test scores	Class size	Discontinuities in class size due to maximum class-size rule	Angrist and Lavy (1999)
College enrollment	Financial aid	Discontinuities in financial aid formula	van der Klaauw (1996)
Health	Heart attack surgery	Proximity to cardiac care centers	McClellan, McNeil and Newhouse (1994)
Crime	Police	Electoral cycles	Levitt (1997)
Employment and Earnings	Length of prison sentence	Randomly assigned federal judges	Kling (1999)
Birth weight	Maternal smoking	State cigarette taxes	Evans and Ringel (1999)

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives, 15*(4), 69–85.

Difference in differences

Difference in differences (DID or DD) is a statistical technique used in the social sciences that attempts to mimic an experimental research design using observational study data, by studying the differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment. It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group. Although it is intended to mitigate the effects of extraneous factors and selection bias, depending on how the treatment group is chosen, this method may still be subject to certain biases (e.g., omitted variable bias).



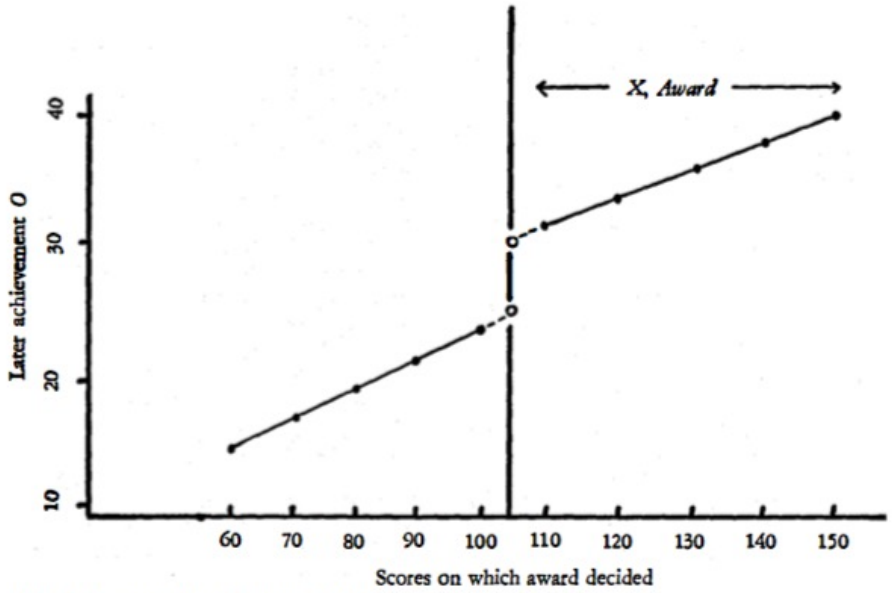
$$Y = B_0 + B_1 Group + B_2 Time + B_3 Group*Time$$

Problem: Assumption that the change in outcomes from pre- to post-intervention in the control group (S) is a good proxy for the (counterfactual) change in untreated potential outcomes in the treated group (P) may not be warranted; choice of treatment/control groups is crucial (an additional trick may be *matching* on observables)…

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? The Quarterly Journal of Economics, 119(1), 249–275.

Regression discontinuity

A regression discontinuity design (RDD) is a quasi-experimental pretest-posttest design that elicits the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned. By comparing observations lying closely on either side of the threshold, it is possible to estimate the average treatment effect in environments in which randomization is unfeasible. RDD was first applied by Donald Thistlethwaite and Donald Campbell to the evaluation of scholarship programs.



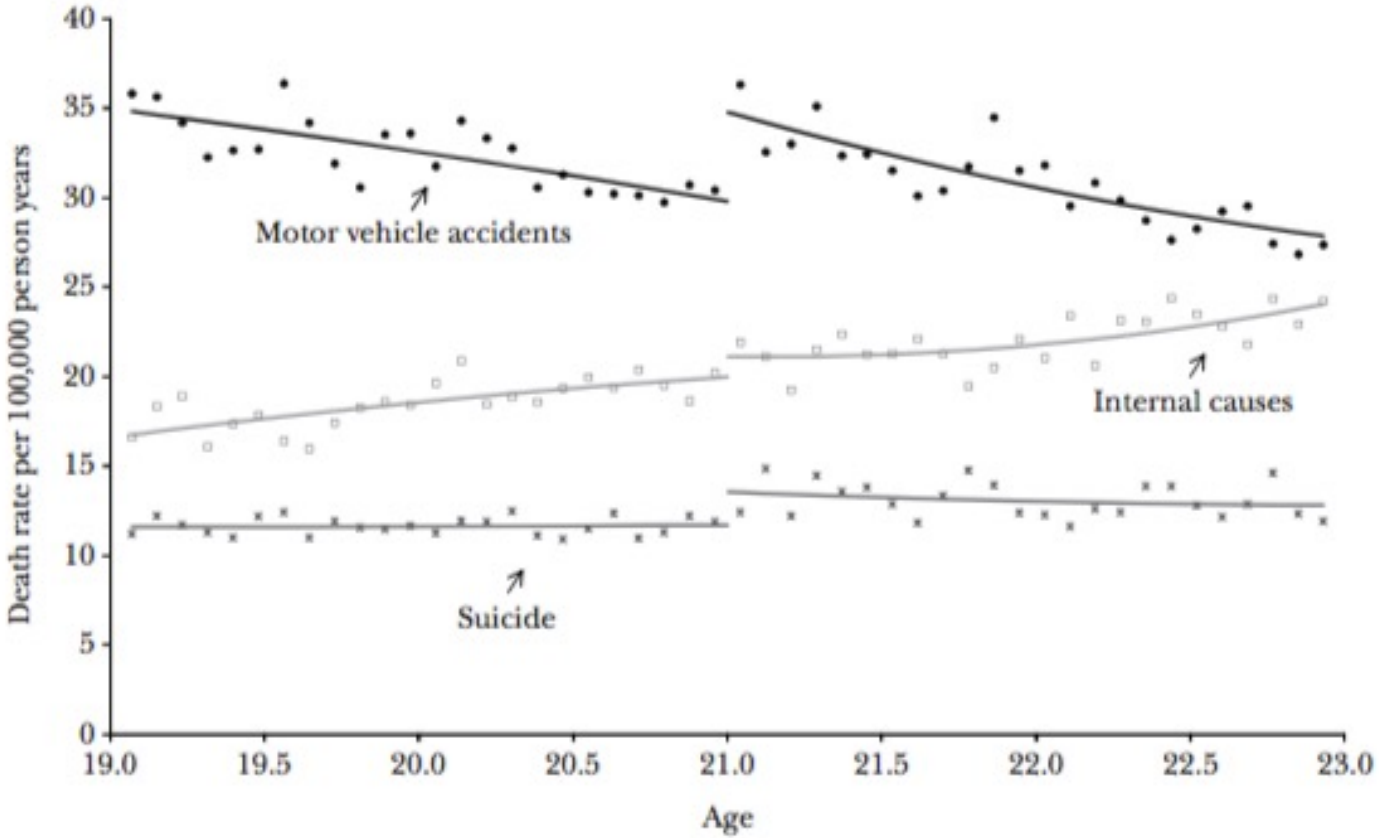
$$Y = B_0 + B_1 \text{Score} + B_2 \text{Award}$$

Problem: Assumption that the individuals just below the cutoff are not systematically different from those just above can be wrong (e.g., individuals just above the threshold could try harder); the estimation may not generalise to observations away from the cutoff (e.g., awards could have different results at different levels of ability).

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature, 48*(2), 281–355. http://doi.org/10.1257/jel.48.2.281

Regression discontinuity

Figure 2
Age Profiles for Death Rates in the United States

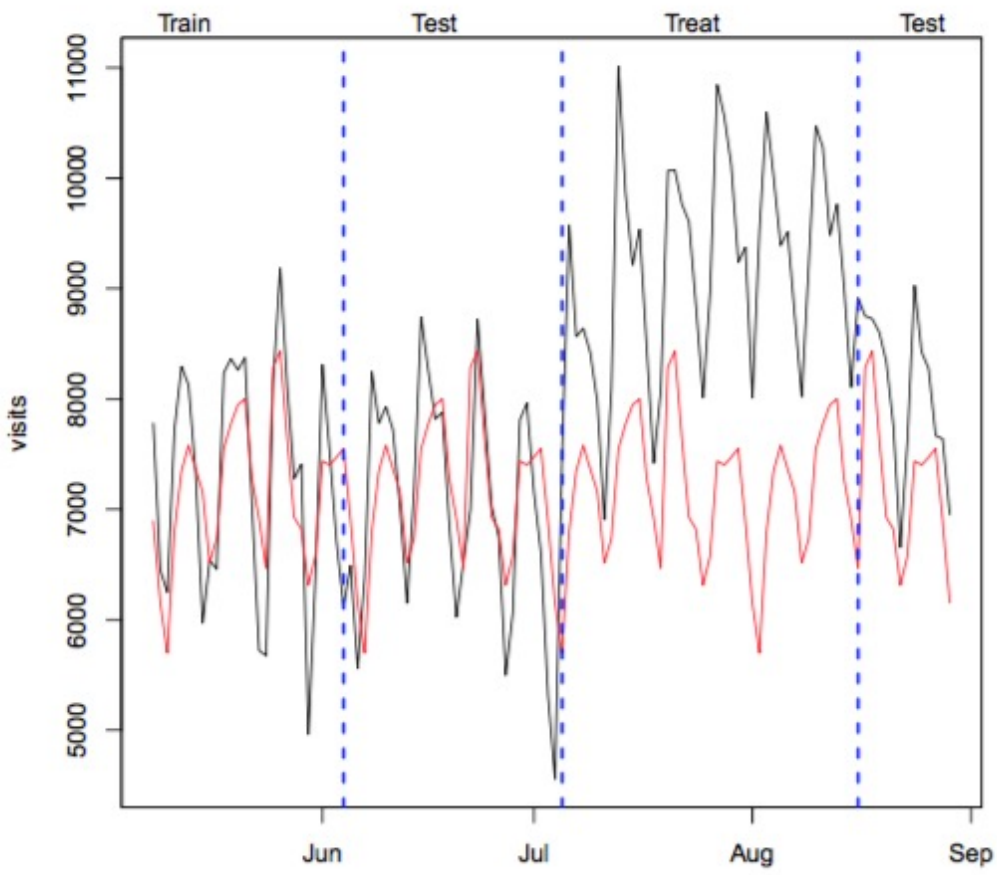


Notes: The death rates are estimated by combining the National Vital Statistics records with population estimates from the U.S. Census.

Carpenter, C., & Dobkin, C. (2011). The Minimum Legal Drinking Age and Public Health. *Journal of Economic Perspectives*, *25*(2), 133–156.

New developments…

Using models as the control group (Train-test-treat-compare)



An online advertiser might ask "if I increase my ad expenditure by some amount, how many extra sales do I generate?"

A predictive statistical model (based on number of "searches" about topics related to the subject matter of the website) is estimated during the training period and its predictive performance is assessed during the test period. The extrapolation of the model during the treat period (red line) serves as a counterfactual. This counterfactual is compared with the actual outcome (black line), and the difference is the estimated treatment effect. When the treatment is ended, the outcome returns to something close to the original level.

Varian, H. R. (2016). Causal inference in economics and marketing. ***Proceedings of the National Academy of Sciences of the United States of America, 113***(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

Summary

"The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment"

There are many types of causal inference analyses that can be (and are) used in the behavioural sciences - in psychology, experiments and multiple regression from observational data are the most commonly used inference methods.

It is helpful to be aware of other methods (e.g., instrumental variables, regression discontinuity, difference in differences) and, more importantly, "the possibility of **creatively** utilizing the idiosyncratic features of any research situation in designing tests of causal hypotheses".