

Why do we randomize?

Fredrik Sävje

Yale University

December 9, 2020

UC Berkeley

STATISTICAL MODELS FOR CAUSATION

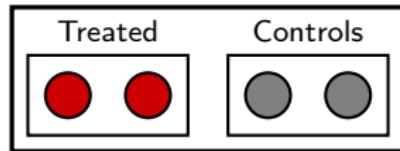
What Inferential Leverage Do They Provide?

DAVID A. FREEDMAN

University of California, Berkeley

The third is the *intention-to-treat estimator*. Although subjects are heterogeneous, the intention-to-treat estimator makes no statistical adjustments for heterogeneity. Instead, randomization is relied upon to balance the treatment and control groups, within the limits of random error. That, after all, is the whole point of doing randomized experiments. Adjustments might in the end bring no additional clarity, a topic considered below.

Randomization avoids systematic errors

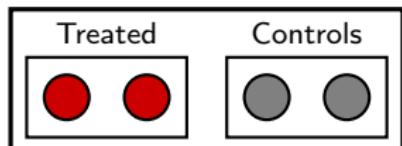


Self-selected treatment assignment

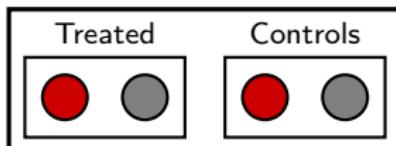
Randomization avoids systematic errors



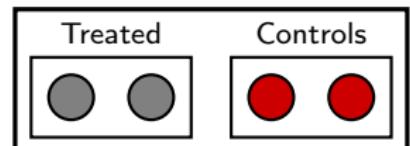
Self-selected treatment assignment



Random assignment 1



Random assignment 2



Random assignment 3



Understanding and misunderstanding randomized controlled trials

Angus Deaton^{a,b,c,*}, Nancy Cartwright^{d,e}

^a Princeton University, USA

^b National Bureau of Economic Research, USA

^c University of Southern California, USA

^d Durham University, England

^e UC San Diego, USA

ARTICLE INFO

Keywords:

RCTs

Balance

Bias

Precision

External validity

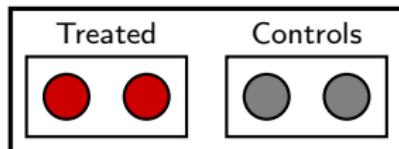
Contrary to frequent claims in the applied literature, randomization does not equalize everything other than the treatment in the treatment and control groups, it does not automatically deliver a precise estimate of the average treatment effect (ATE), and it does not relieve us of the need to think about (observed or unobserved) covariates.

ABSTRACT

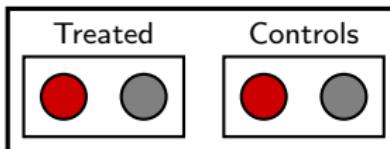
Randomized Controlled Trials (RCTs) are increasingly popular in the social sciences, not only in medicine. We argue that the lay public, and sometimes researchers, put too much trust in RCTs over other methods of investigation. Contrary to frequent claims in the applied literature, randomization does *not* equalize everything other than the treatment in the treatment and control groups; it does not automatically deliver a precise estimate

of the average treatment effect (ATE), and it does not relieve us of the need to think about (observed or unobserved) covariates. External validity is required to extend the results to other groups, including any population to which the trial sample belongs, or to any individual, including an individual in the trial. Demanding ‘external validity’ is unhelpful because it expects too much of an RCT while undervaluing its potential contribution. RCTs do indeed require minimal assumptions and can operate with little prior knowledge. This is an advantage when persuading distrustful audiences, but it is a disadvantage for cumulative scientific progress, where prior knowledge should be built upon, not discarded. RCTs can play a role in building scientific knowledge and useful predictions but they can only do so as part of a cumulative program, combining with other methods, including conceptual and theoretical development, to discover not ‘what works’, but ‘why things work’.

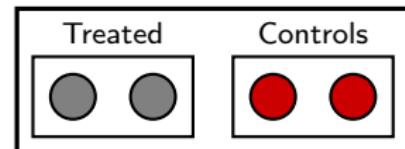
Randomization does not avoid unsystematic errors



Random assignment 1

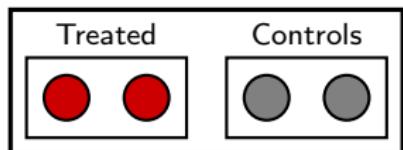


Random assignment 2



Random assignment 3

Randomization does not avoid unsystematic errors



Random assignment 1



Random assignment 2

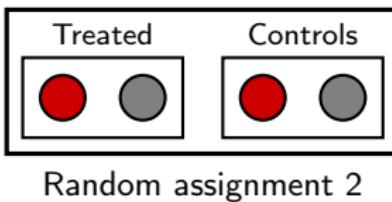


Random assignment 3

D&C: "Is this really a good assignment?"

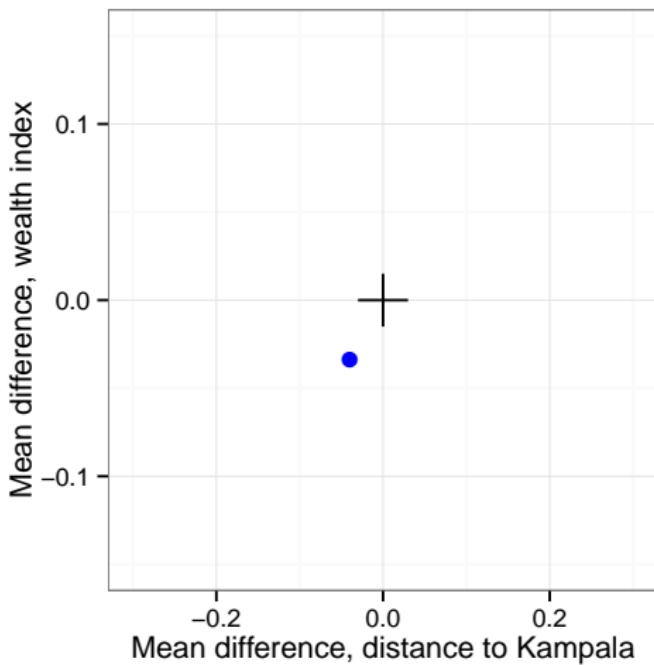


Randomization does not avoid unsystematic errors

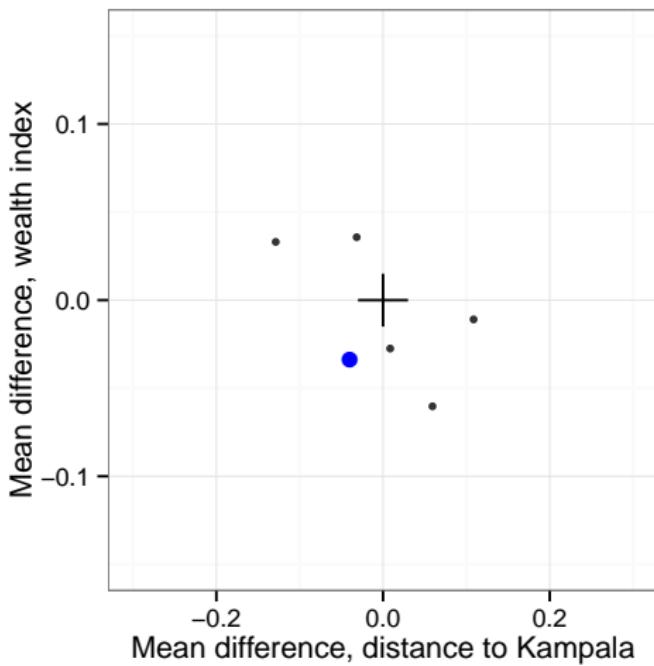


D&C: “Make your experiments balanced!”

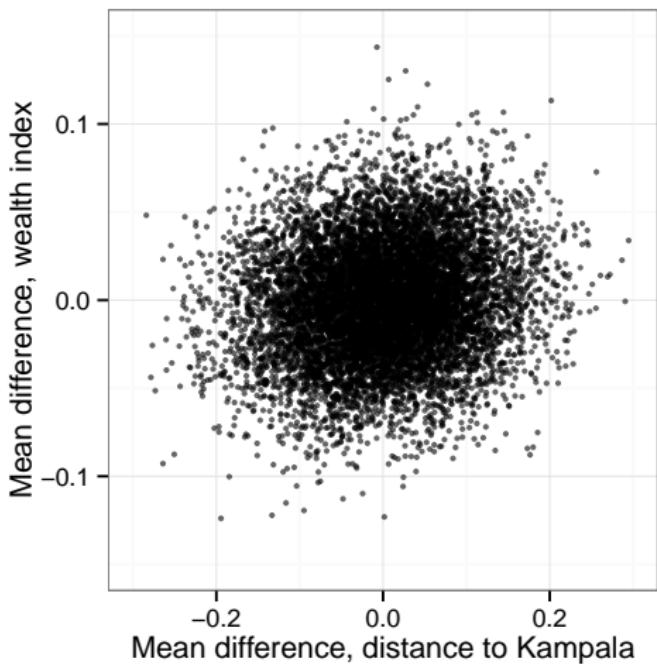
Example: Grossman, Humphreys & Sacramone-Lutz (2014)



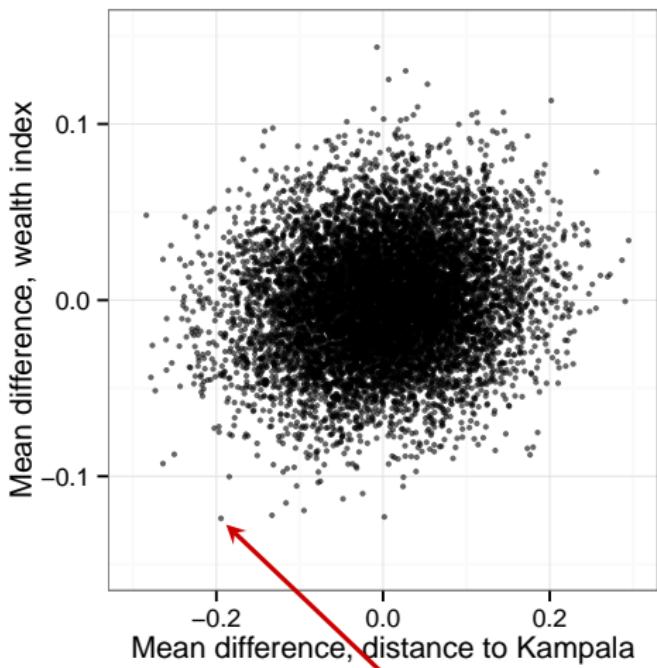
Example: Grossman, Humphreys & Sacramone-Lutz (2014)



Example: Grossman, Humphreys & Sacramone-Lutz (2014)

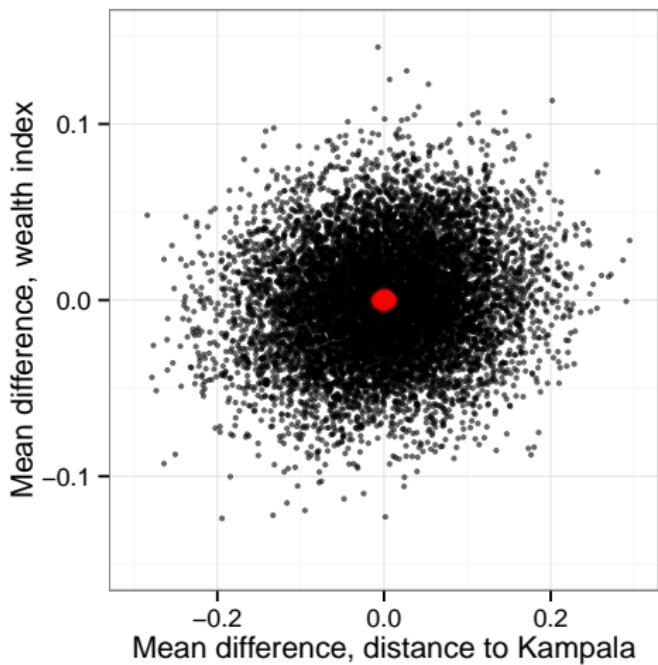


Example: Grossman, Humphreys & Sacramone-Lutz (2014)

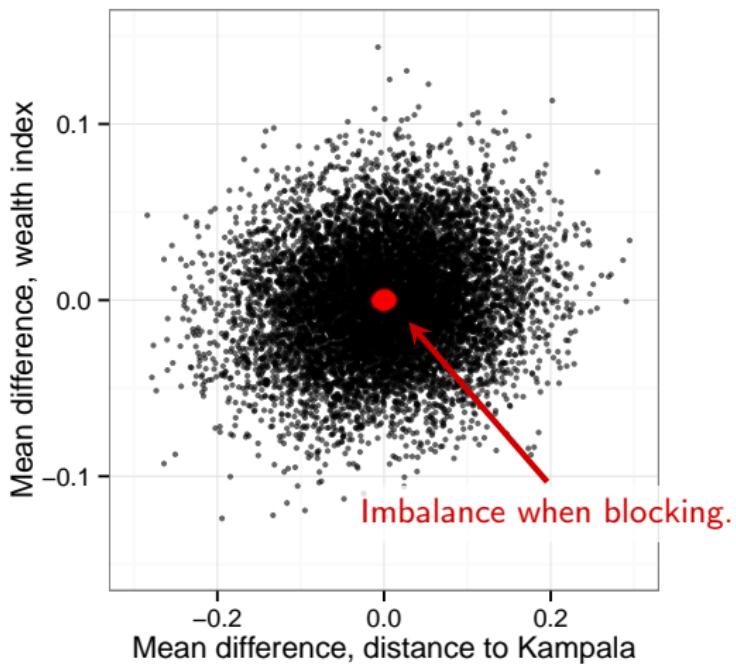


D&C: "Is this really a good assignment?"

Example: Grossman, Humphreys & Sacramone-Lutz (2014)



Example: Grossman, Humphreys & Sacramone-Lutz (2014)



Misunderstandings between experimentalists and observationalists about causal inference

Kosuke Imai,

Princeton University, USA

Gary King

Harvard University, Cambridge, USA

and Elizabeth A. Stuart

Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

490 *K. Imai, G. King and E. A. Stuart*

founders are known and so can be adjusted for exactly, even if all covariates are not available. Thus, except in very small samples, **blocking on pretreatment variables followed by random treatment assignment cannot be worse than randomization alone**. Blocking on variables related to the outcome is of course more effective in increasing statistical efficiency than blocking on irrelevant variables, and so it pays to choose the variables to block carefully. But choosing not to block on a relevant pretreatment variable before randomization, that is feasible to use, is not justified.

5.2. Assumptions

Experimentalists and observationalists often make assumptions about unobserved processes on the basis of prior evidence or theory. At worst, when the question is judged to be sufficiently

OPTIMAL STRATIFICATION IN RANDOMIZED EXPERIMENTS

THOMAS BARRIOS

DEPARTMENT OF ECONOMICS, HARVARD UNIVERSITY

The Power of Optimization Over Randomization in Designing Experiments Involving Small Samples

Dimitris Bertsimas

Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, bertsim@mit.edu

Mac Johnson

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, mac.johnson@sloan.mit.edu

Nathan Kallus

Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, kallus@mit.edu

Optimal multivariate matching before randomization

ROBERT GREEVY*

Department of Statistics, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340, USA

BO LU

Center for Statistical Sciences, Department of Community Health, Brown University School of Medicine, Providence, RI 02912, USA,

JEFFREY H. SILBER

Department of Pediatrics, University of Pennsylvania, School of Medicine, 3535 Market Street, STE 1029, Philadelphia, PA 19104-3309, USA

PAUL ROSENBAUM

Department of Statistics, The Wharton School, University of Pennsylvania, 473 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340, USA

The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation

Kosuke Imai, Gary King and Clayton Nall

OPTIMAL STRATIFICATION IN RANDOMIZED EXPERIMENTS

In Pursuit of Balance: Randomization in Practice in Development Field Experiments[†]

THOMAS BARRIOS

DEPARTMENT OF ECONOMICS, HARVARD UNIVERSITY

By MIRIAM BRUHN AND DAVID MCKENZIE[‡]

The Power of Optimization over Randomization in Designing Experiments Involvin

RERANDOMIZATION TO IMPROVE COVARIATE BALANCE IN EXPERIMENTS[†]

BY KARI LOCK MORGAN AND DONALD B. RUBIN

Dimitris Bertsimas

Operations Research Center, Massachusetts Institute of Technology, Cambridge,

Mac Johnson

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Ma

Nathan Kallus

Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, kallus@mit.edu

Optimal multivariate matching before randomization

ROBERT GREEVY*

Wharton School, University of Pennsylvania, 400 Jon M. Huntsman
Valnut Street, Philadelphia, PA 19104-6340, USA

BO LU

...z, Department of Community Health, Brown University School of
Medicine, Providence, RI 02912, USA,

JEFFREY H. SILBER

Department of Pediatrics, University of Pennsylvania, School of Medicine, 3535 Market Street,
STE 1029, Philadelphia, PA 19104-3309, USA

PAUL ROSENBAUM

Department of Statistics, The Wharton School, University of Pennsylvania, 473 Jon M. Huntsman

The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Uni Health Insurance Evaluation

Kosuke Imai, Gary King and Clayton Nall

Improving massive experiments with threshold blocking

Michael J. Higgins^a, Fredrik Sävje^b, and Jasjeet S. Sekhon^{c,d,†}

SEQUENTIAL TREATMENT ASSIGNMENT WITH BALANCING FOR PROGNOSTIC FACTORS IN THE CONTROLLED CLINICAL TRIAL

STUART J. FOOCOK

Statistical Laboratory, SUNY at Buffalo, Amherst, New York 14260, U.S.A.

RICHARD SIMON

National Cancer Institute, Bethesda, Maryland 20204

Multivariate Continuous Blocking to Improve Political Science Experiments

OPT

ization in Practice ...
Development Field Experimentation

By MIRIAM BRUHN AND DAVID N.

The Power of Optimization Over Random Designing Experiments Involvin

Dimitris Bertsimas

Operations Research Center, Massachusetts Institute of Technology, Cambridge,

Mac Johnson

Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Ma

RER

Nathan Kallus

Cornell University, New York, USA

IN

Ryan T. Moore

THOMAS BAKRIOS

DEPT OF ELECTRICAL & COMPUTER ENGINEERING

Optimal *a priori* balance in the design of controlled experiments

DUANE V. RILOCK MORGAN AND DONALD B. RUBIN

Asymptotic theory of rerandomization in treatment-control experiments

Optimal multivariate matching before randomization

ROBERT GREEVY*

Xinran Li^a, Peng Ding^b, and Donald B. Rubin^{a,1}

Forcing a sequential experiment to be balanced

^aPennsylvania, 400 Jon M. Huntsman
A 19104-6340, USA

Nearly random designs with greatly improved balance

BY A. M. KRIEGER

Department of Statistics, The Wharton School of the University of Pennsylvania, 3730 Walnut Street,
Philadelphia, Pennsylvania 19104, U.S.A.

krieger@wharton.upenn.edu

^bHealth, Brown University School of
12, USA,

D. AZRIEL

Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology,
Technion City, Haifa 3200003, Israel
davidazr@technion.ac.il

^cMedicine, 3535 Market Street,
4-3309, USA

The Essential Role of Cluster-Randomized Application to the Medical Health Insurance Evals

Kosuke Imai, Gary King and Clayton Nall

^dPennsylvania, 473 Jon M. Huntsman

AND A. KAPELNER

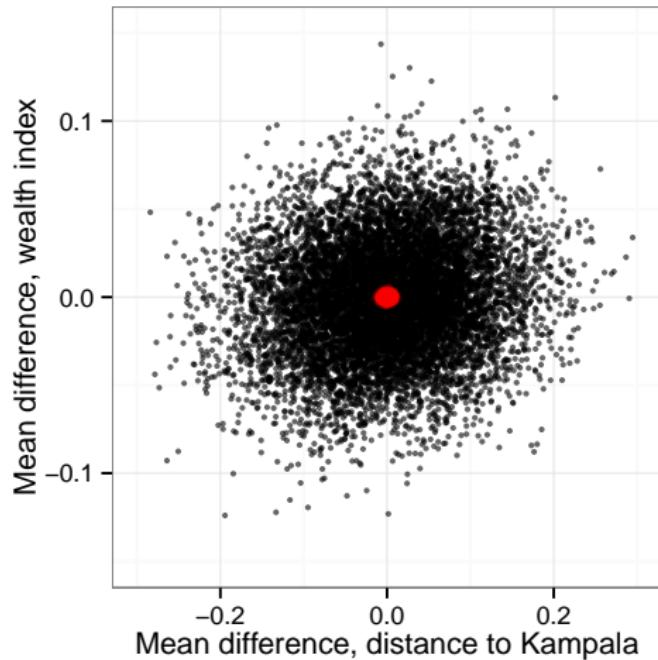
Department of Mathematics, Queens College, City University of New York, 65-30 Kissena Blvd,
Queens, New York 11367, U.S.A.
kapelner@qc.cuny.edu

ents with

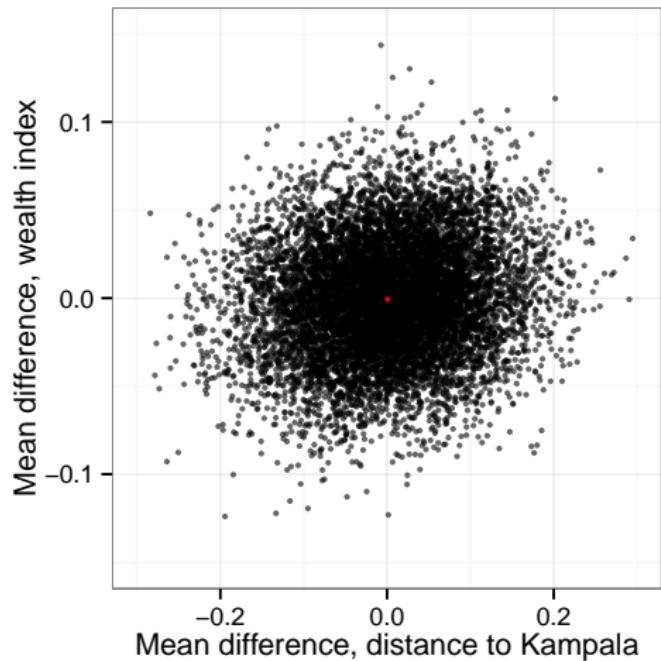
If balance makes experiments better, then...

Maximize covariate balance?

Maximize balance?



Maximize balance?



But if the goal is to maximize balance, then...

Why randomize?

Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead

Maximilian Kasy

*Department of Economics, Harvard University, 1805 Cambridge Street,
Cambridge, MA 02138, USA
e-mail: maximilankasy@fas.harvard.edu (corresponding author)*

Suppose that an experimenter has collected a sample as well as baseline information about the units in the sample. How should she allocate treatments to the units in this sample? We argue that the answer does not involve randomization if we think of experimental design as a statistical decision problem. If, for instance, the experimenter is interested in estimating the average treatment effect and evaluates an estimate in terms of the squared error, then she should minimize the expected mean squared error (MSE) through choice of a treatment assignment. We provide explicit expressions for the expected MSE that lead to easily implementable procedures for experimental design.

1 Introduction

Experiments, and in particular randomized experiments, are the conceptual reference point that gives empirical content to the notion of causality. In recent years, actual randomized experiments have become increasingly popular elements of the methodological toolbox in a wide range of social science disciplines. Examples from the recent political science literature abound. Blattman

Today's talk

How should I assign treatments in my experiments?

- Why should we randomize?
- How should we randomize?
- How can we randomize in practice?

Today's talk

How should I assign treatments in my experiments?

- Why should we randomize?
- How should we randomize?
- How can we randomize in practice?

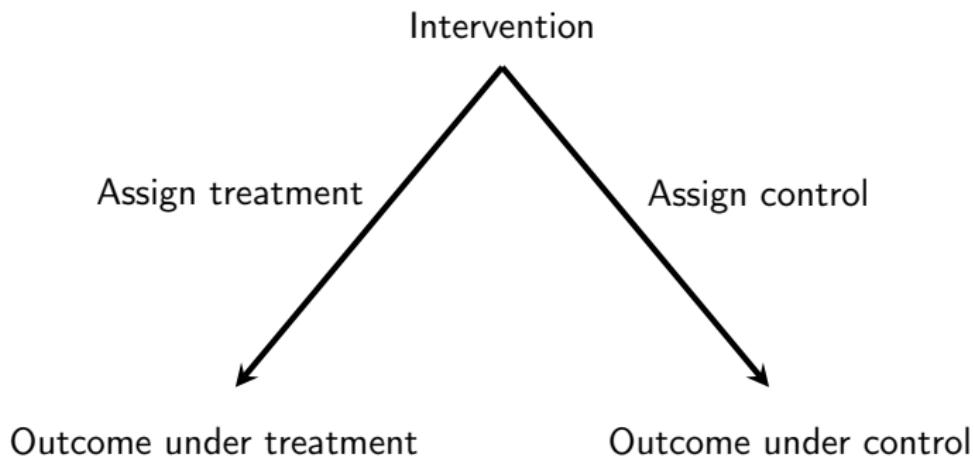
Building a general theory of experiments

- Understand potential outcomes as directions
- Understand experimental designs as ellipses

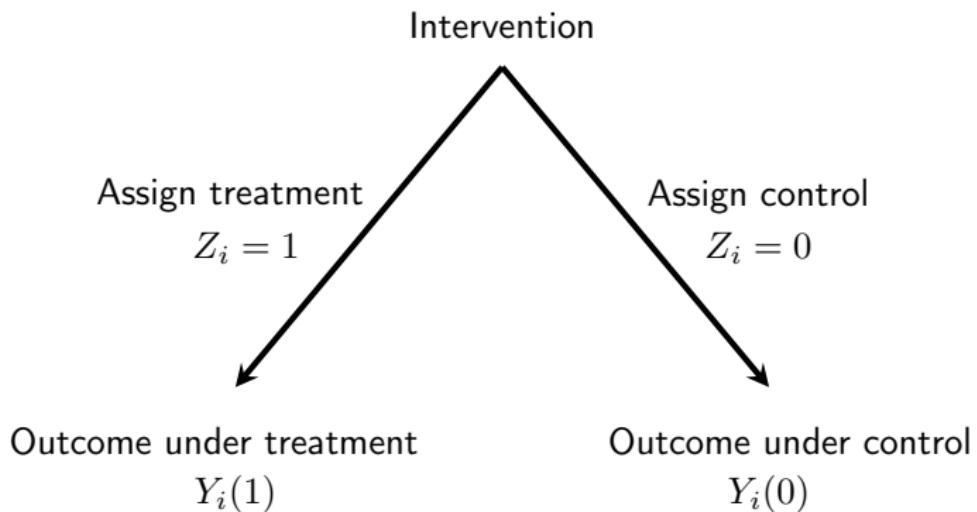
Building a general theory of experiments

- Understand potential outcomes as directions
- Understand experimental designs as ellipses

Potential outcomes



Potential outcomes



Two important aspects of the potential outcomes

Treatment effect

$$\tau_i = Y_i(1) - Y_i(0)$$

Used to define the average treatment effect.

Mean potential outcome

$$\mu_i = \frac{Y_i(1) + Y_i(0)}{2}$$

Helps us understand why and how to randomize.

Two important aspects of the potential outcomes

Treatment effect

$$\tau_i = Y_i(1) - Y_i(0)$$

Used to define the average treatment effect.

Mean potential outcome

$$\mu_i = \frac{Y_i(1) + Y_i(0)}{2}$$

Helps us understand why and how to randomize.

Understanding the potential outcomes as directions

$$Y_1(1) = 2$$

$$Y_1(0) = 0$$

$$\mu_1 = (2 + 0)/2 = 1$$

$$Y_2(1) = 4$$

$$Y_2(0) = -1$$

$$\mu_2 = (4 - 1)/2 = 1.5$$

Understanding the potential outcomes as directions

$$Y_1(1) = 2$$

$$Y_1(0) = 0$$

$$\mu_1 = (2 + 0)/2 = 1$$

$$Y_2(1) = 4$$

$$Y_2(0) = -1$$

$$\mu_2 = (4 - 1)/2 = 1.5$$

Understanding the potential outcomes as directions

$$Y_1(1) = 2$$

$$Y_1(0) = 0$$

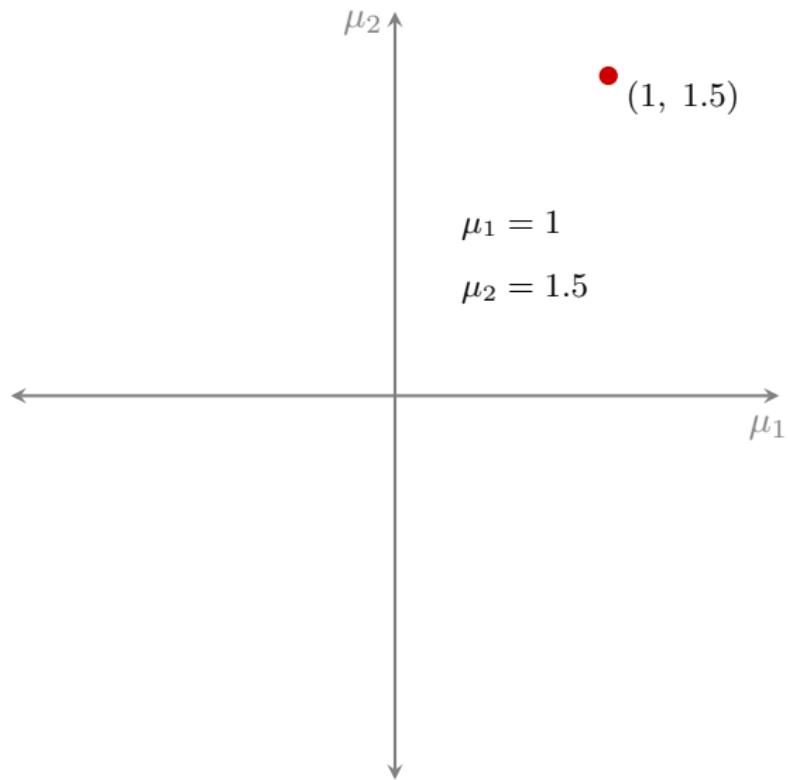
$$\mu_1 = 1$$

$$Y_2(1) = 4$$

$$Y_2(0) = -1$$

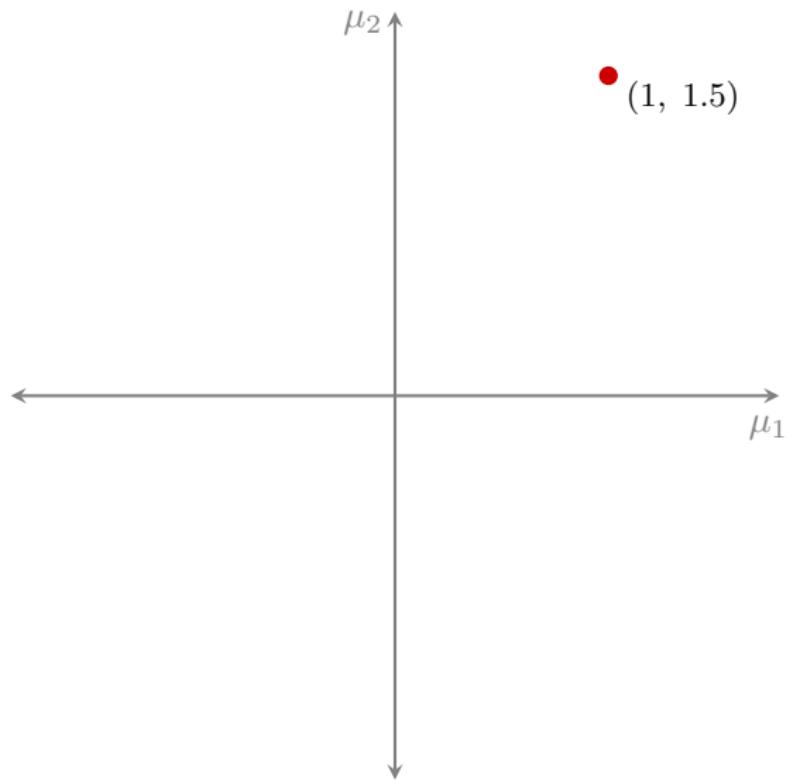
$$\mu_2 = 1.5$$

Understanding the potential outcomes as directions



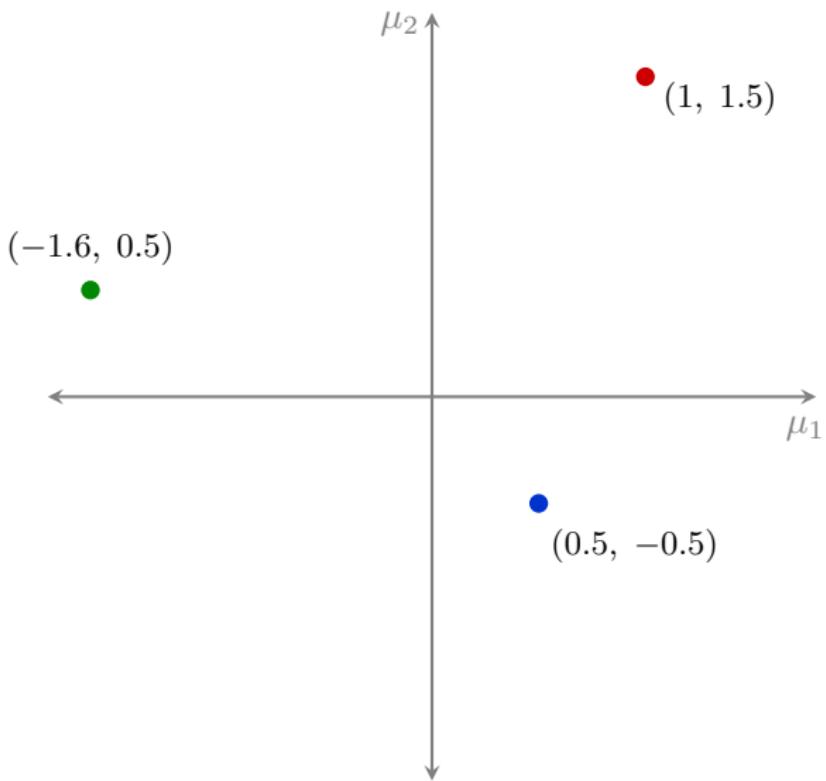
Details

Understanding the potential outcomes as directions



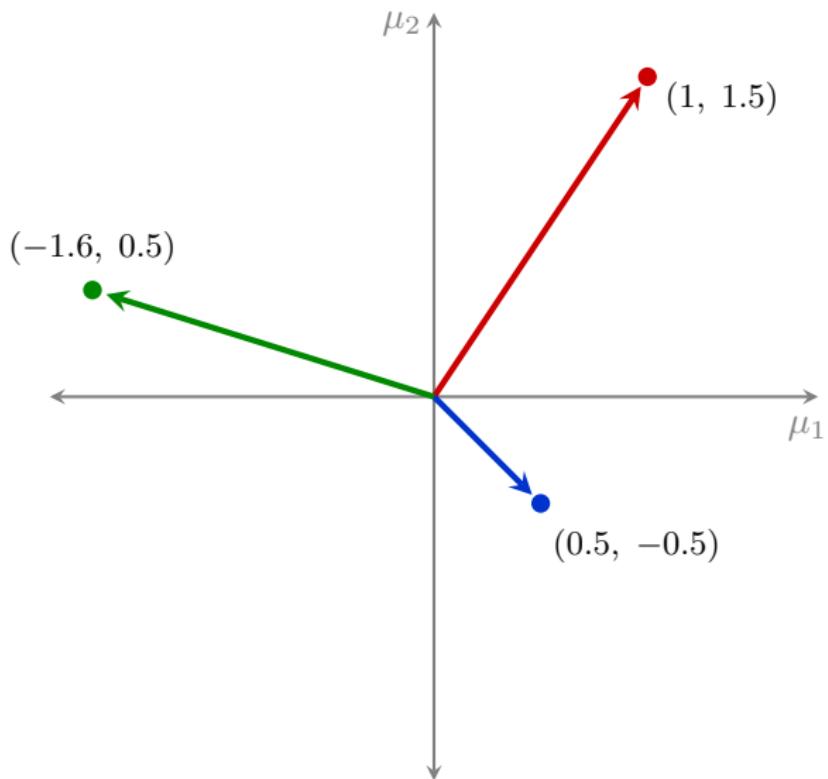
Details

Understanding the potential outcomes as directions



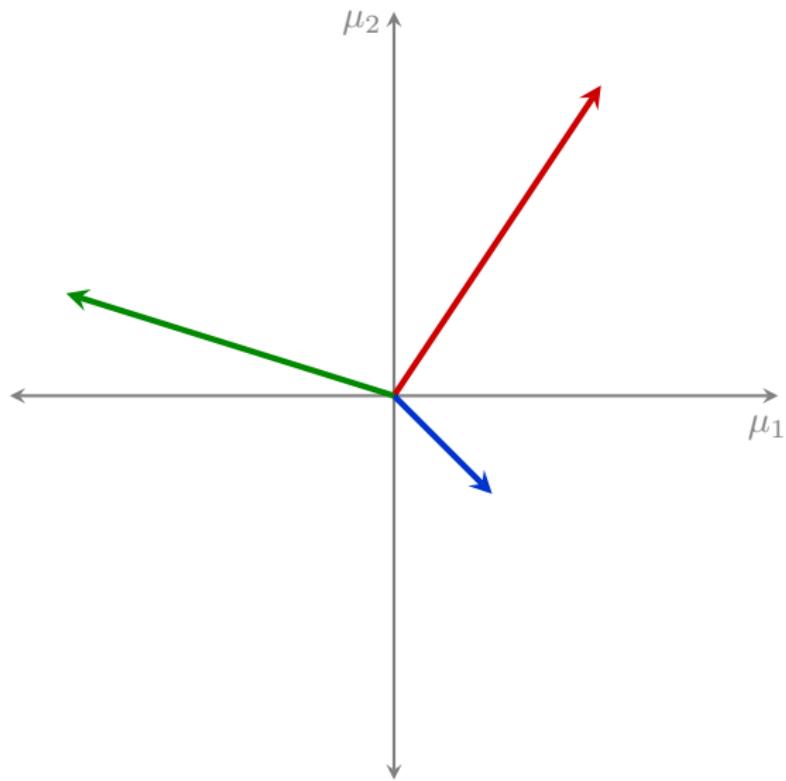
Details

Understanding the potential outcomes as directions



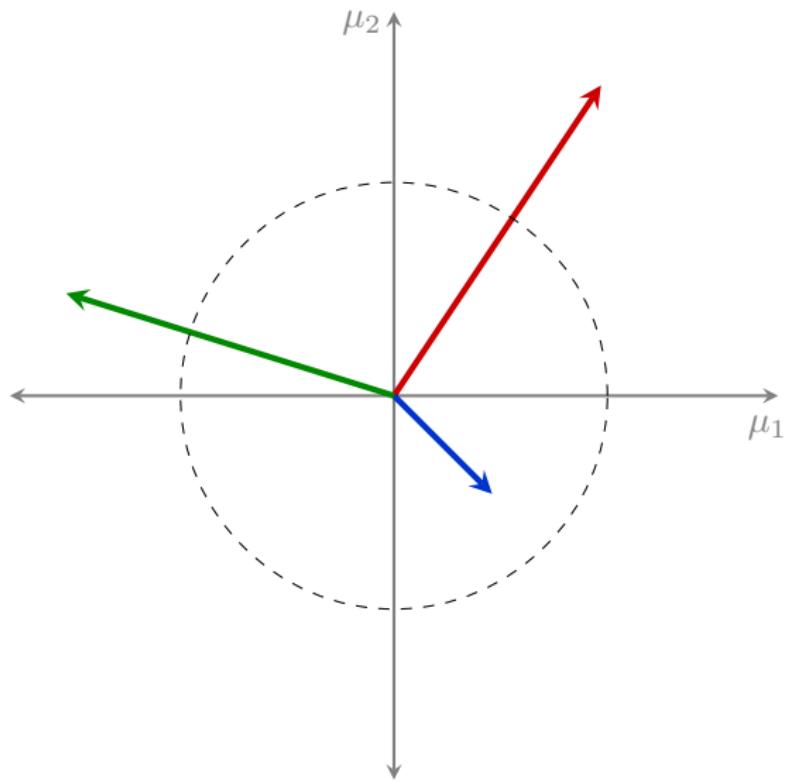
Details

Understanding the potential outcomes as directions



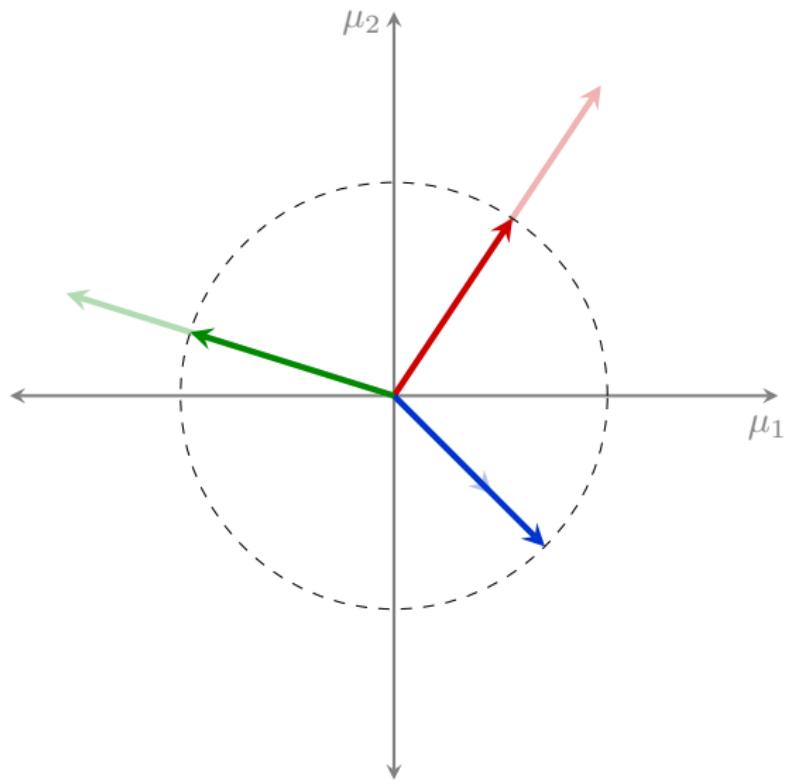
Details

Understanding the potential outcomes as directions



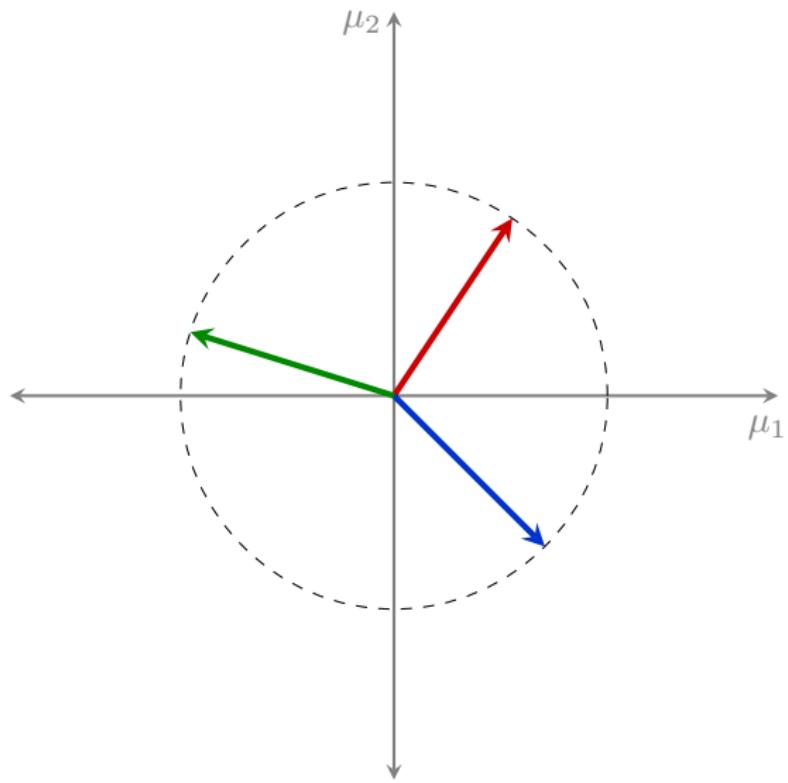
Details

Understanding the potential outcomes as directions



Details

Understanding the potential outcomes as directions



Details

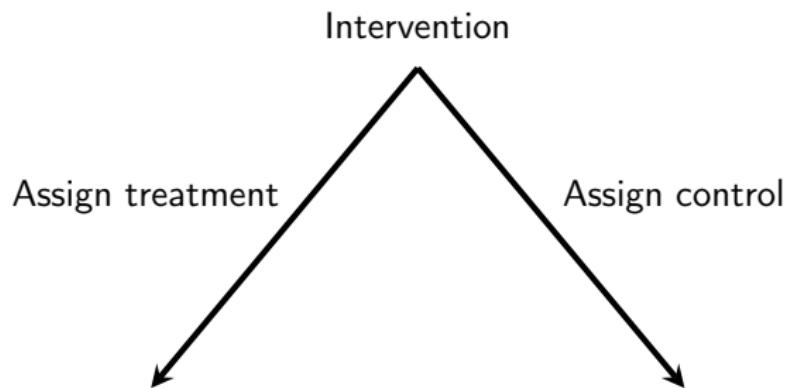
Building a general theory of experiments

- Understand potential outcomes as directions ✓
- Understand experimental designs as ellipses

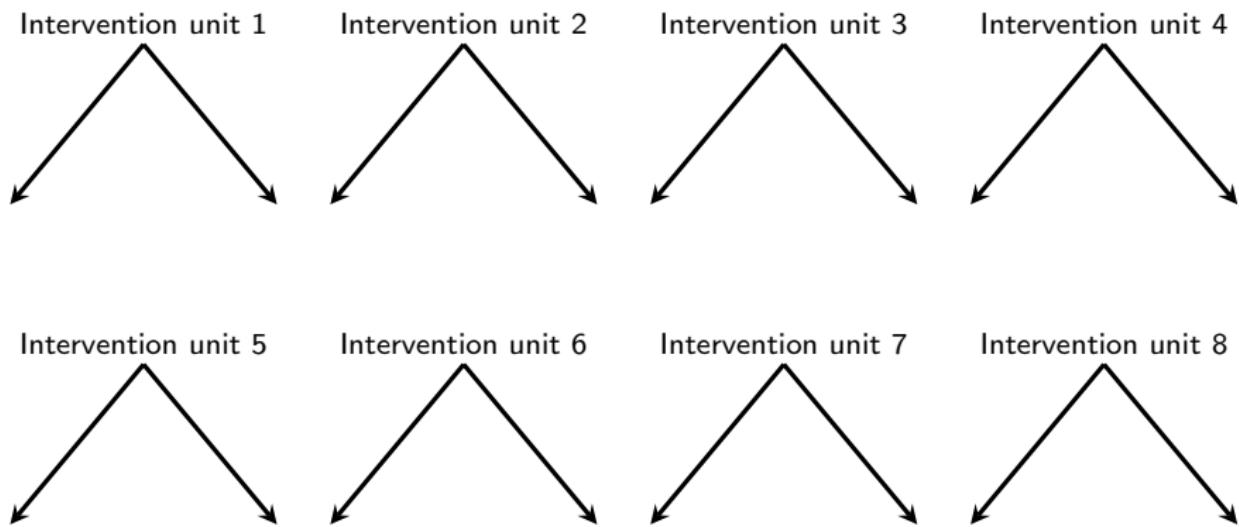
Building a general theory of experiments

- Understand potential outcomes as directions ✓
- Understand experimental designs as ellipses

(Statistical) experimental design



(Statistical) experimental design



(Statistical) experimental design

Intervention unit 1

Intervention unit 2

Intervention unit 3

Intervention unit 4

Assignment vector: $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$.

Intervention unit 5

Intervention unit 6

Intervention unit 7

Intervention unit 8

The probability distribution of \mathbf{Z} is the *experimental design*.

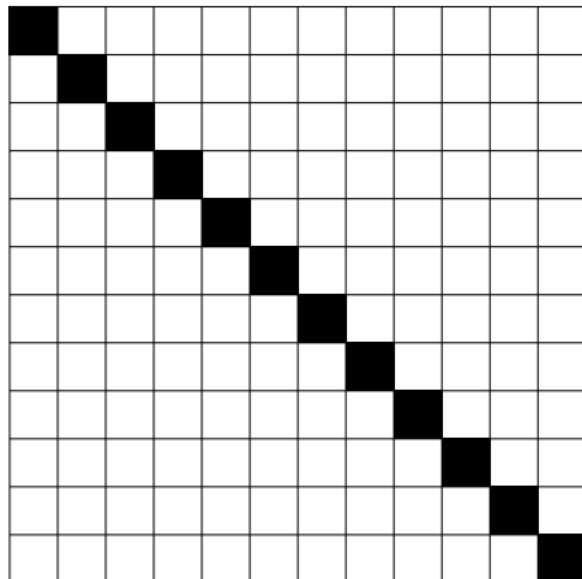
Fingerprints of experimental designs

$$\text{Cov}(\mathbf{Z}) =$$

A blank 10x10 grid for drawing or plotting. The grid consists of 100 equal-sized squares arranged in a single column and ten rows.

Fingerprints of experimental designs

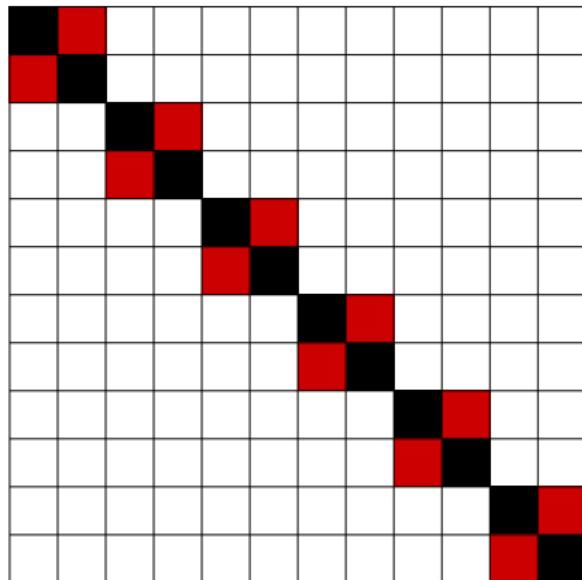
$$\text{Cov}(\mathbf{Z}) =$$



Independent assignment

Fingerprints of experimental designs

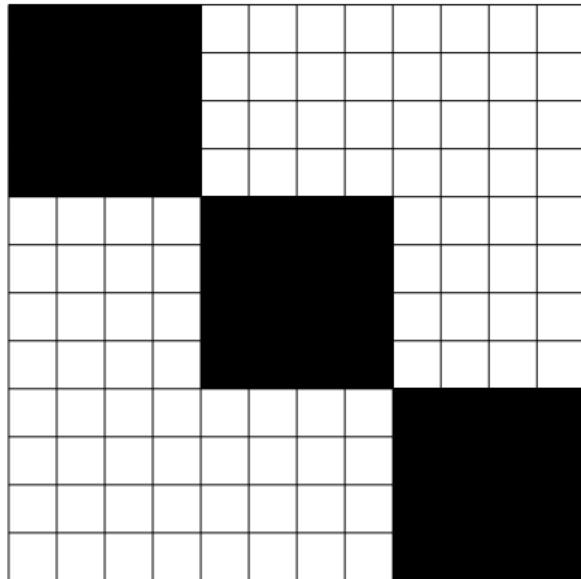
$$\text{Cov}(\mathbf{Z}) =$$



Matched pair design

Fingerprints of experimental designs

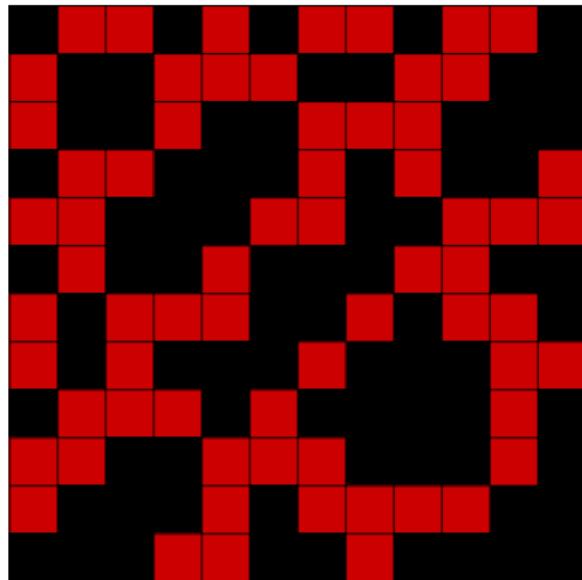
$$\text{Cov}(\mathbf{Z}) =$$



Cluster assignment

Fingerprints of experimental designs

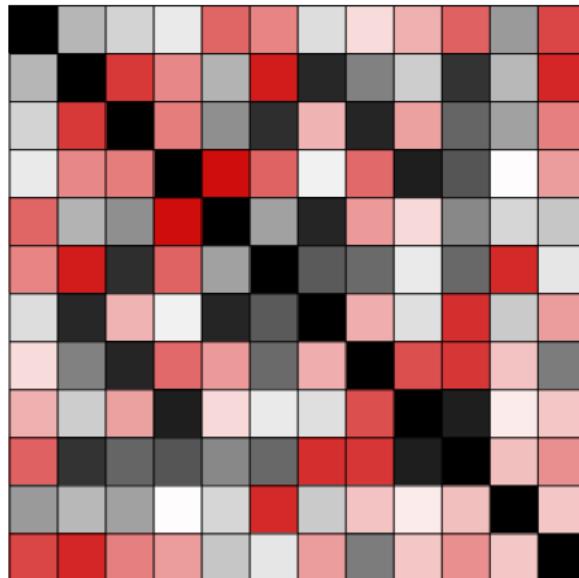
$$\text{Cov}(\mathbf{Z}) =$$



Deterministic assignment (e.g., Kasy, 2016)

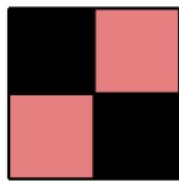
Fingerprints of experimental designs

$$\text{Cov}(\mathbf{Z}) =$$



Rerandomization

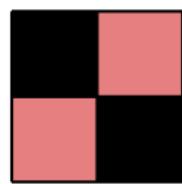
Understanding experimental designs as ellipses



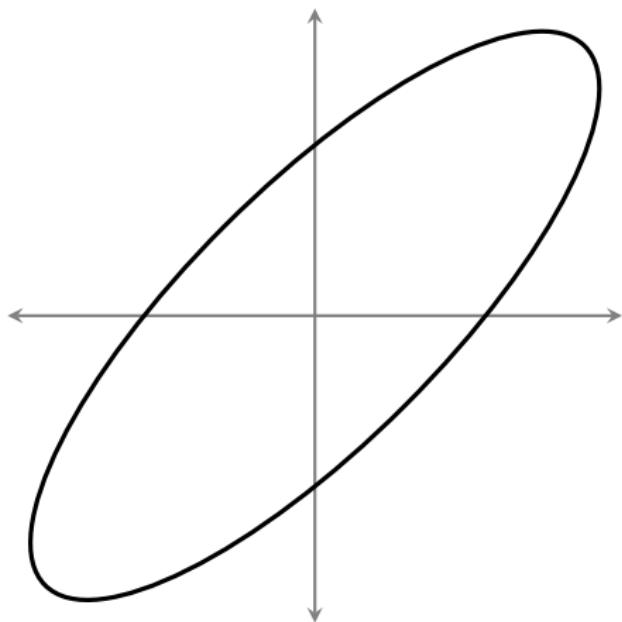
Details

$n = 3$

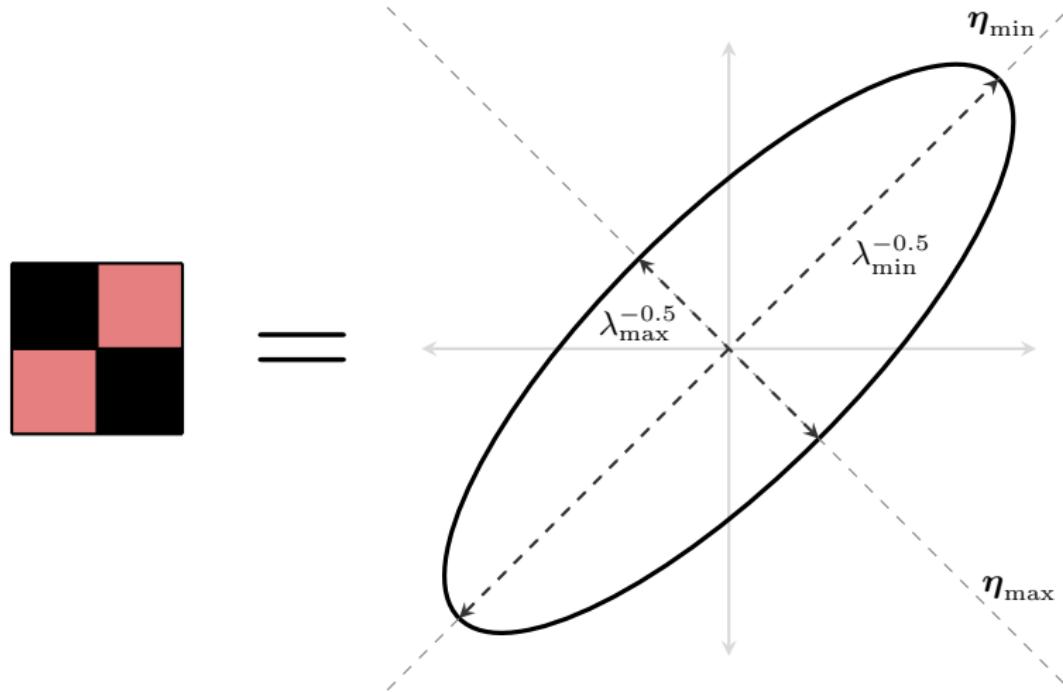
Understanding experimental designs as ellipses



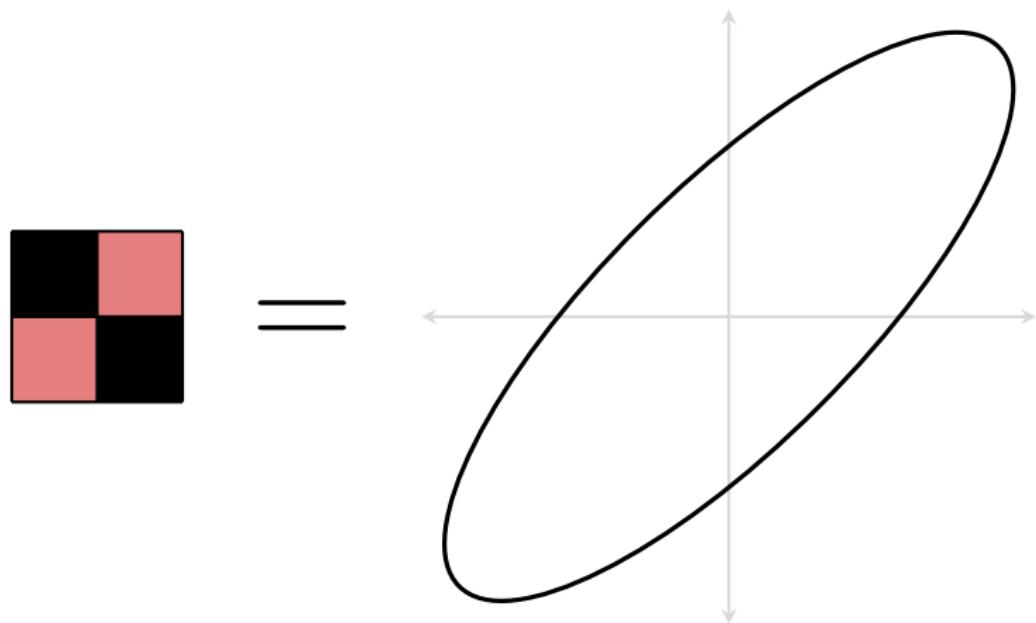
=



Understanding experimental designs as ellipses



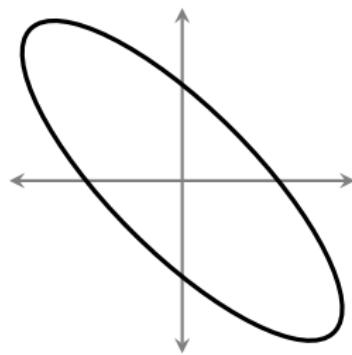
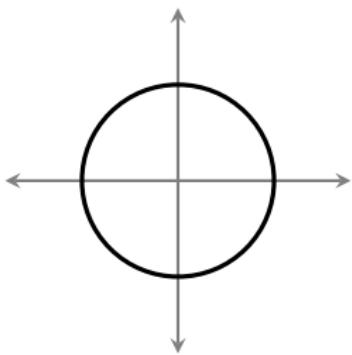
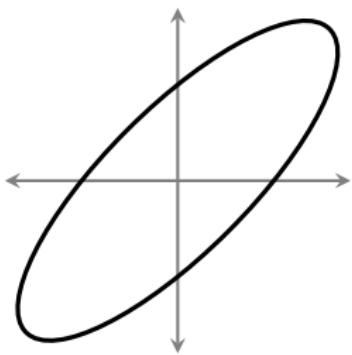
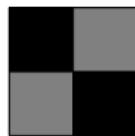
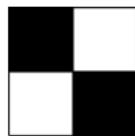
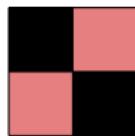
Understanding experimental designs as ellipses



Details

$n = 3$

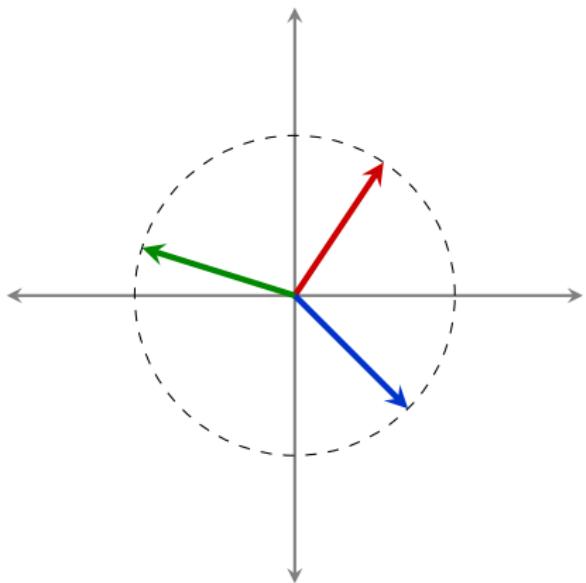
Understanding experimental designs as ellipses



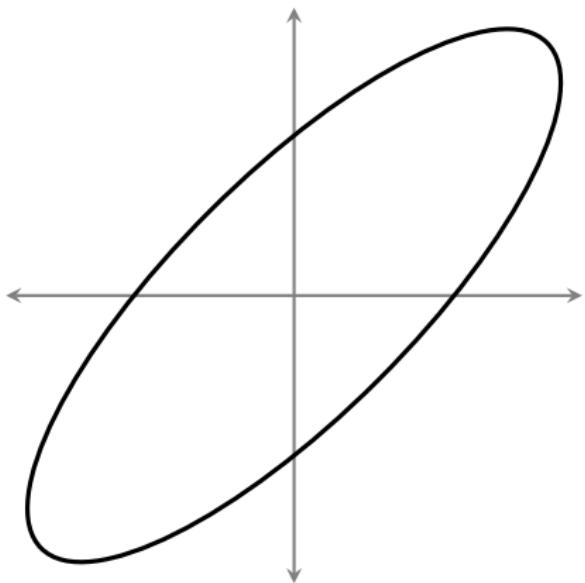
Building a general theory of experiments

- Understand potential outcomes as directions ✓
- Understand experimental designs as ellipses ✓

Connecting potential outcomes and designs

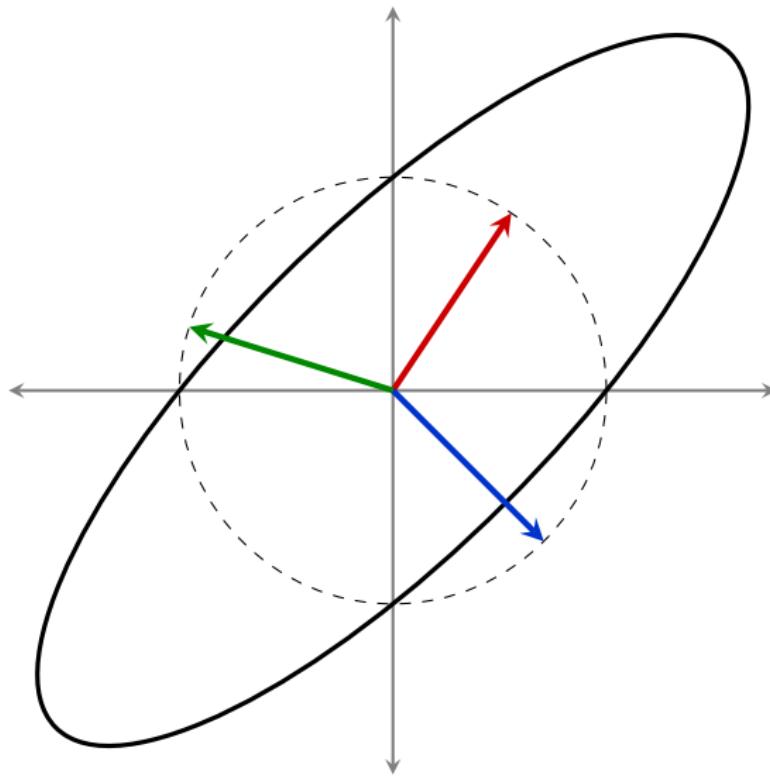


Potential outcomes



Experimental design

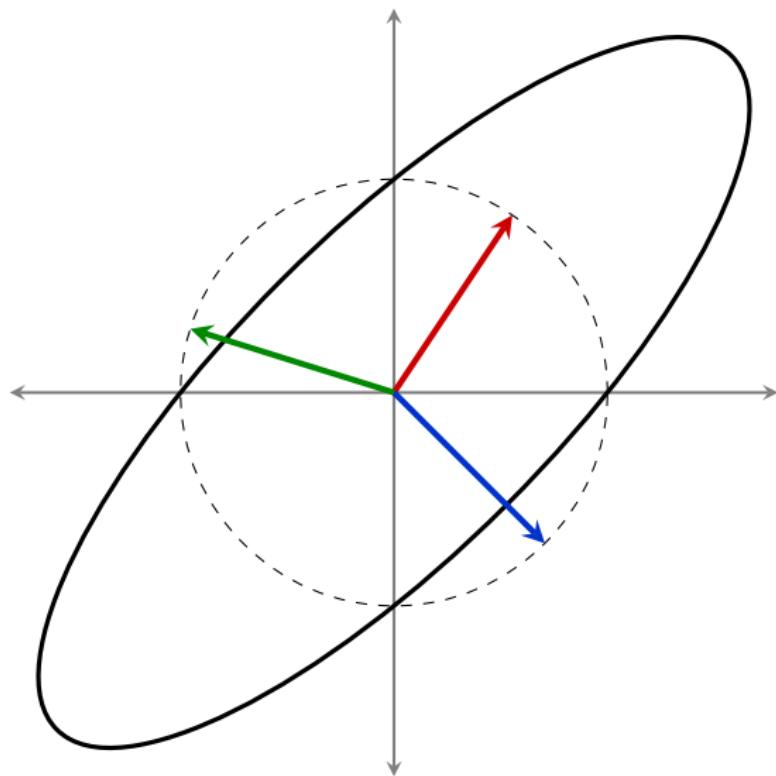
Connecting potential outcomes and designs



Theorem

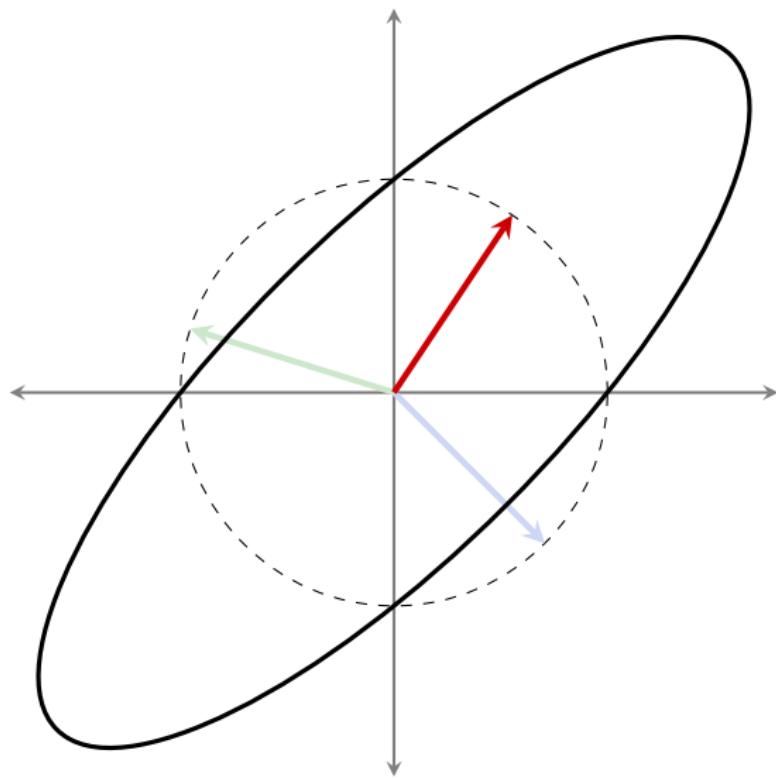
An experimental design performs well if its ellipse is aligned with the direction of the potential outcomes.

A design performs well if it's aligned with the outcomes



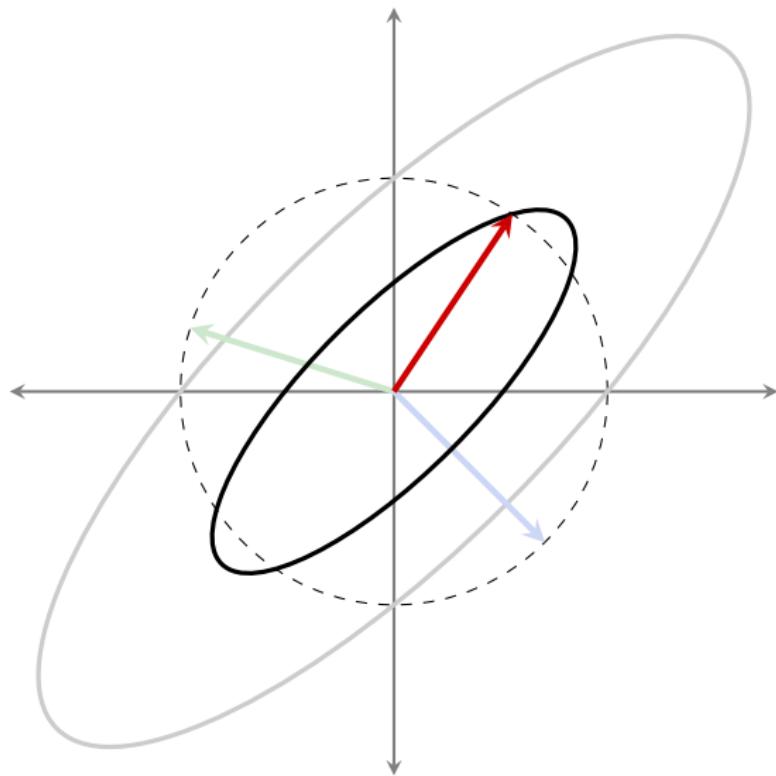
Details

A design performs well if it's aligned with the outcomes



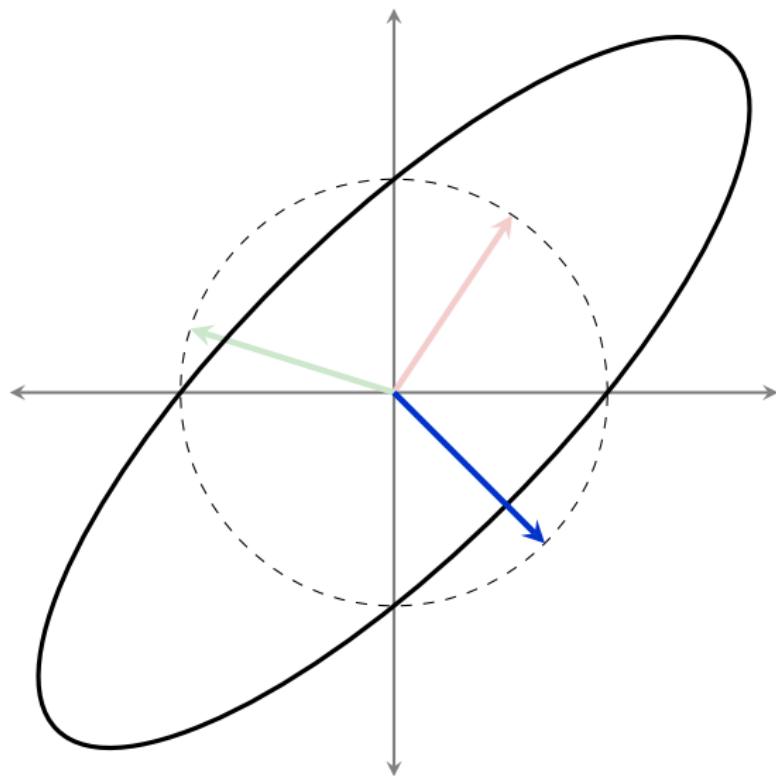
Details

A design performs well if it's aligned with the outcomes



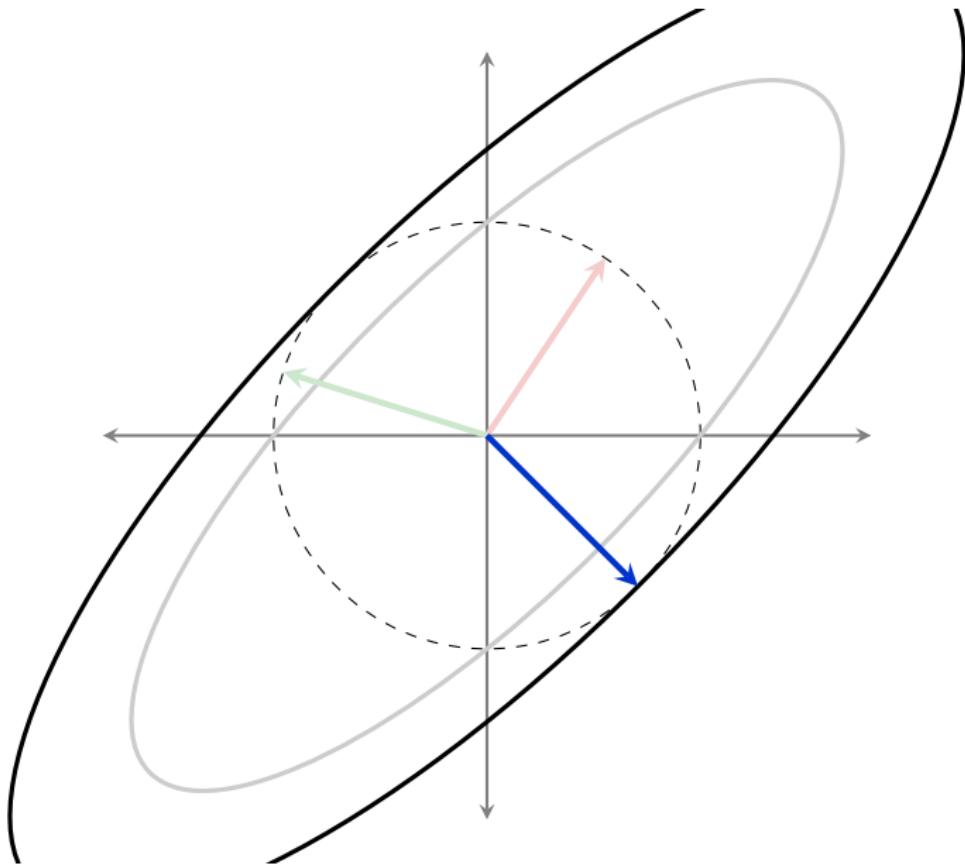
Details

A design performs well if it's aligned with the outcomes



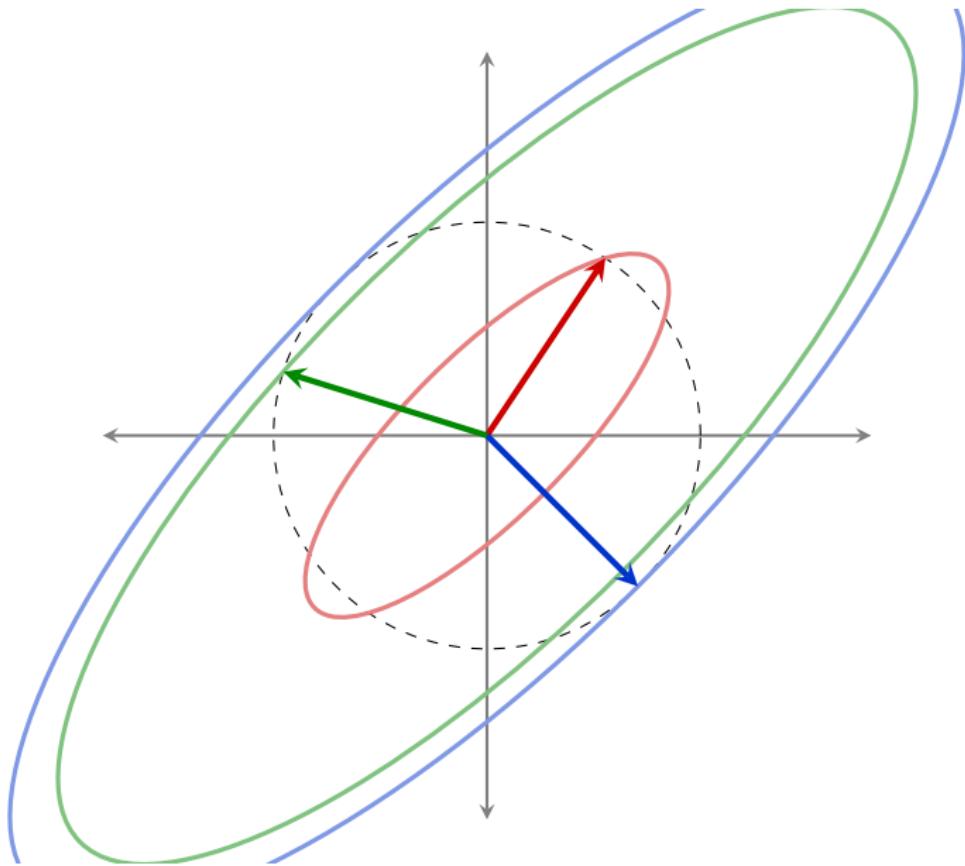
Details

A design performs well if it's aligned with the outcomes



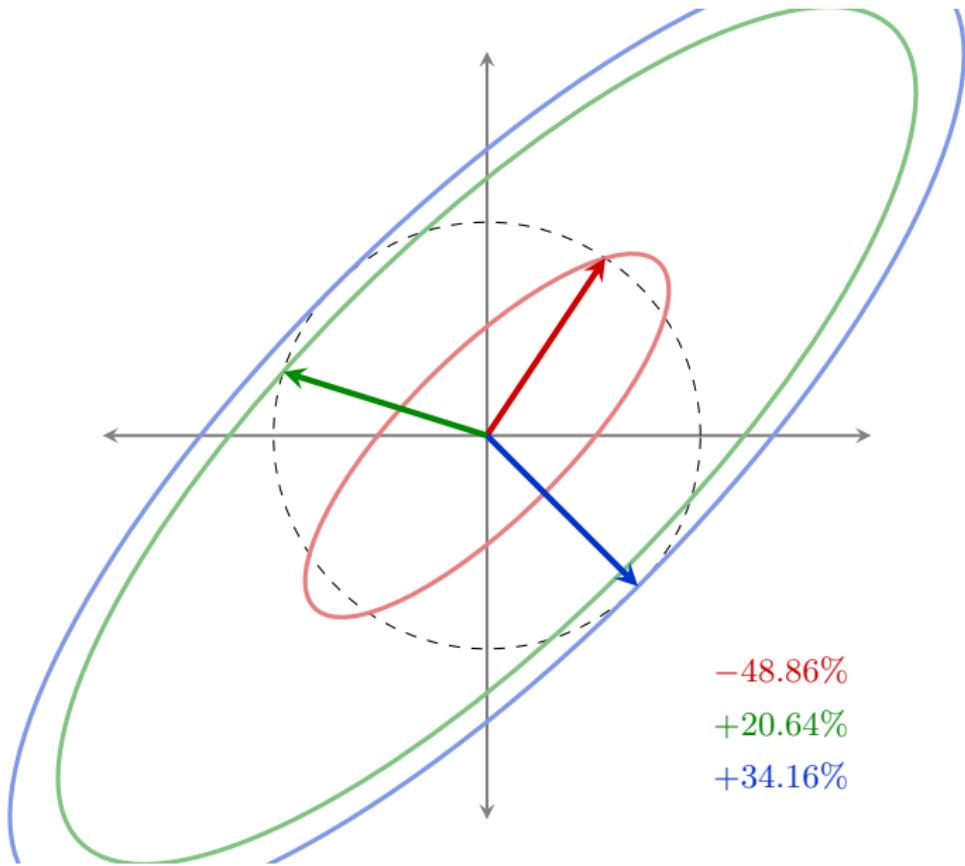
Details

A design performs well if it's aligned with the outcomes



Details

A design performs well if it's aligned with the outcomes



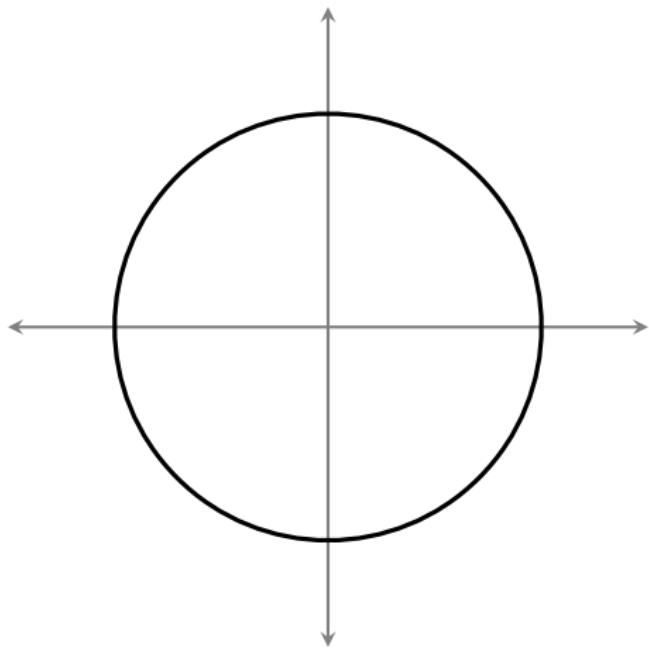
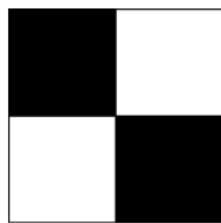
Details

Theorem

Independent assignment performs equally well in all potential outcome directions.

Theorem

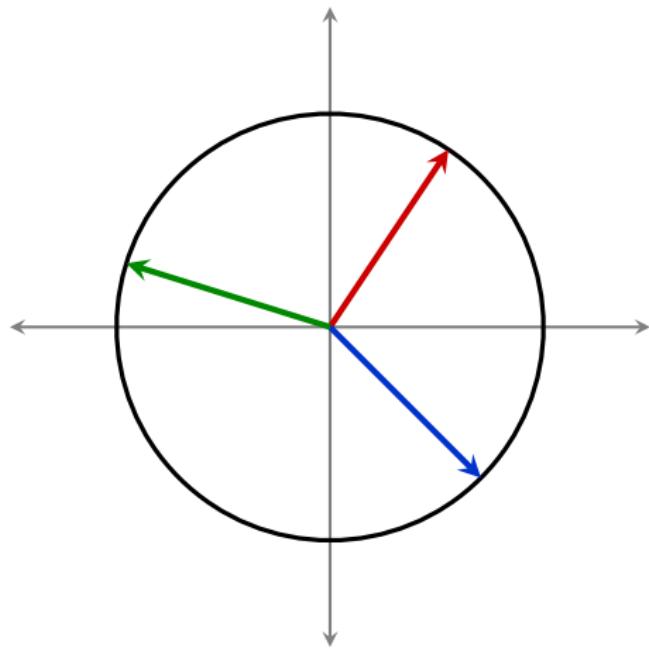
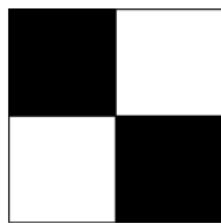
Independent assignment performs equally well in all potential outcome directions.



Details

Theorem

Independent assignment performs equally well in all potential outcome directions.



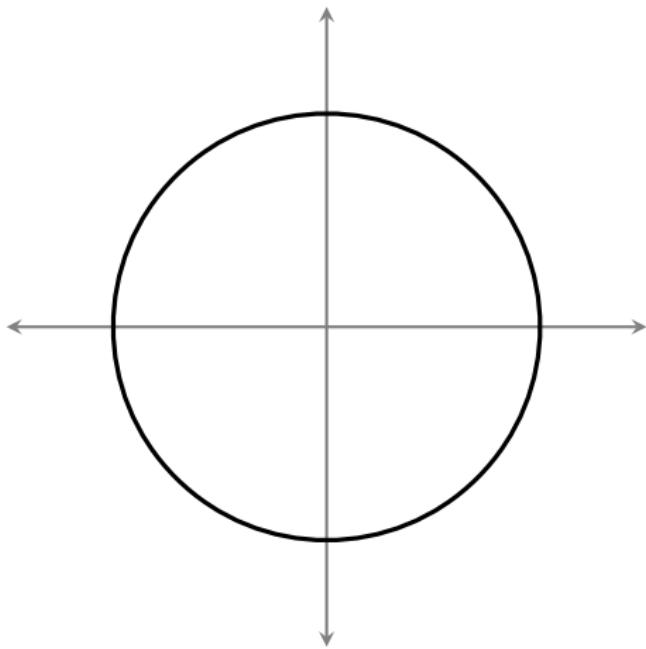
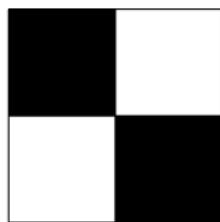
Details

Theorem

We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.

Theorem

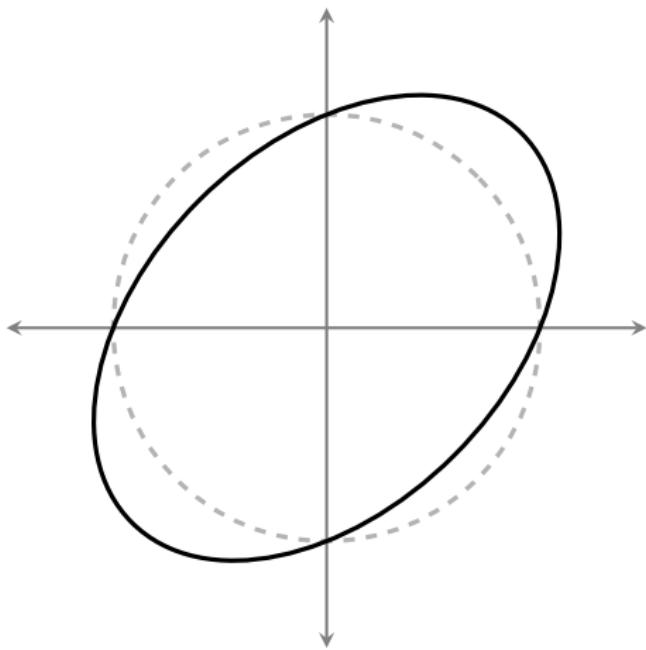
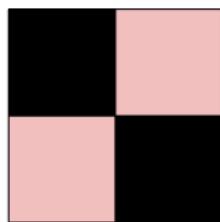
We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.



Details

Theorem

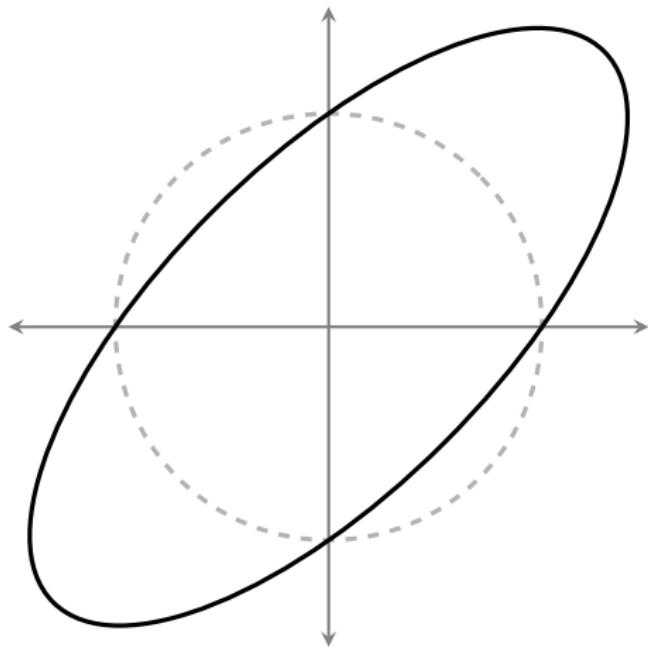
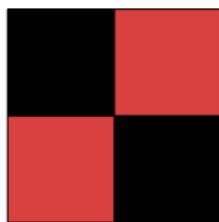
We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.



Details

Theorem

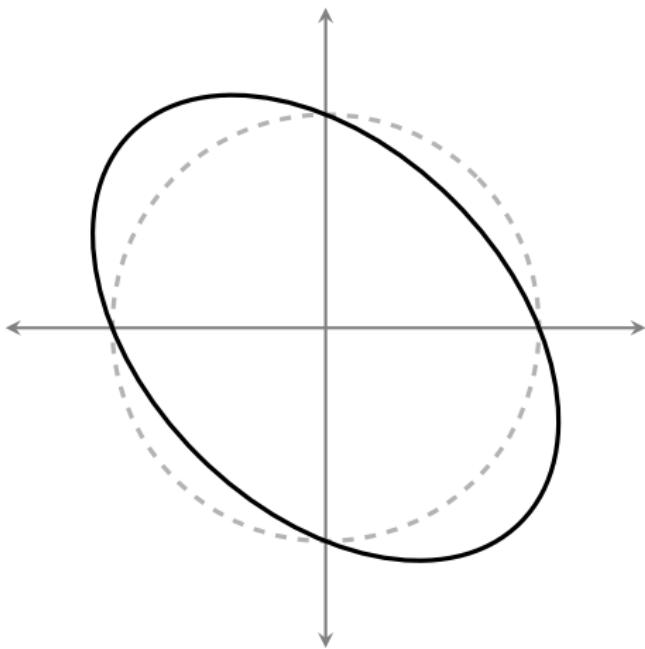
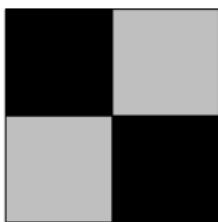
We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.



Details

Theorem

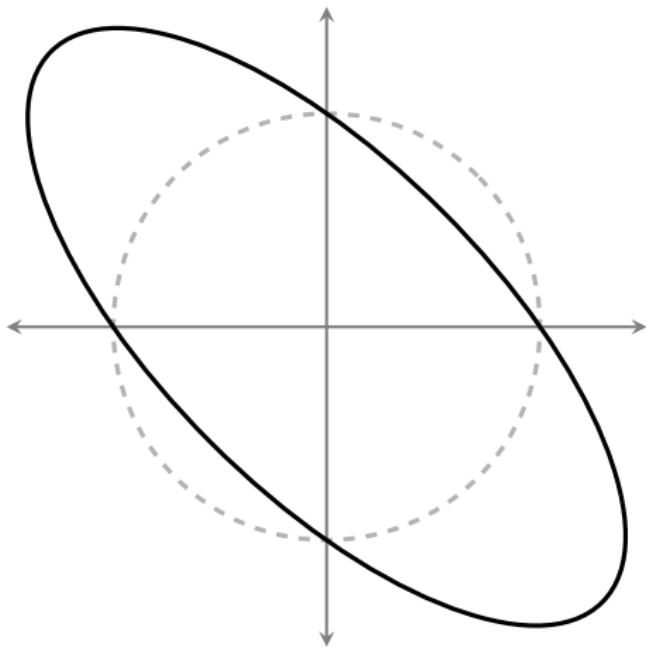
We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.



Details

Theorem

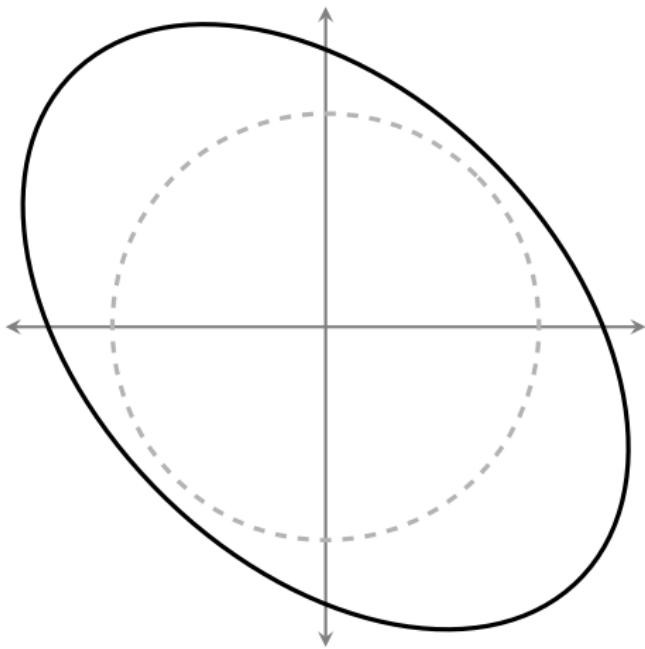
We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.



Details

Theorem

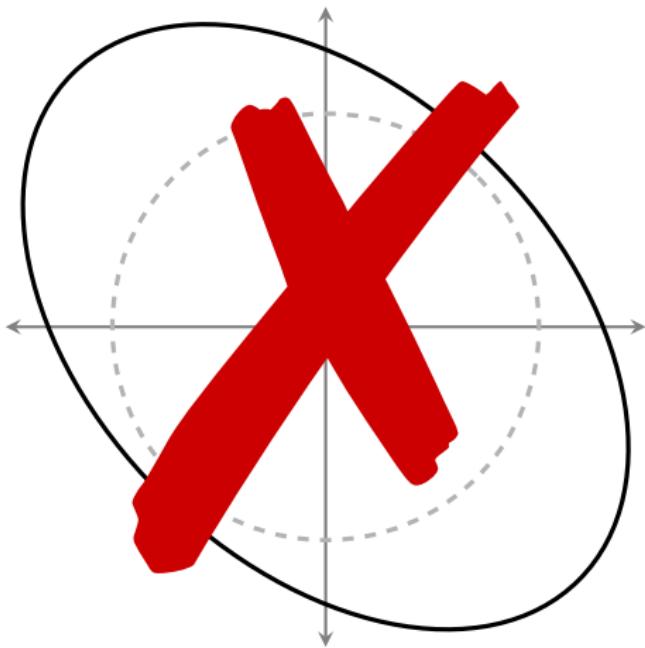
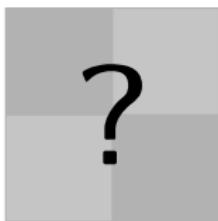
We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.



Details

Theorem

We can change the design by “squeezing” the ellipse, but it’s not possible to change its overall size.



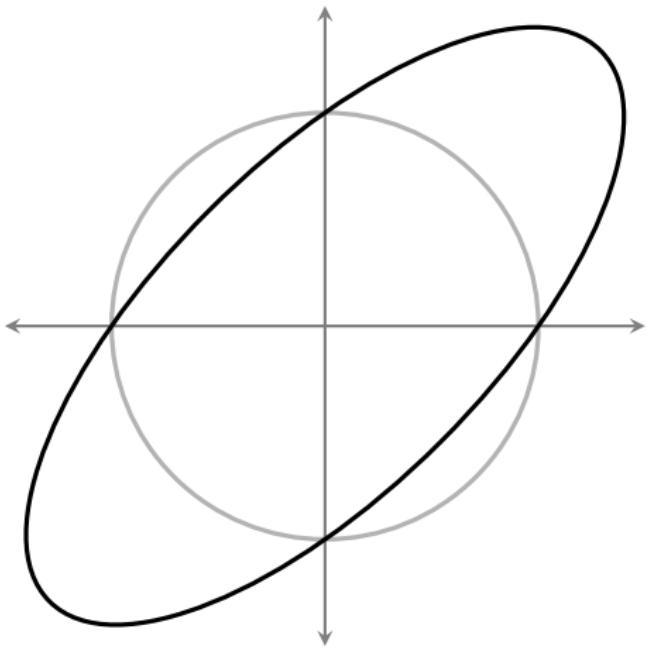
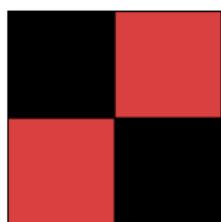
Details

Corollary

It's impossible to make a design perform better than independent assignment for all potential outcomes.

Corollary

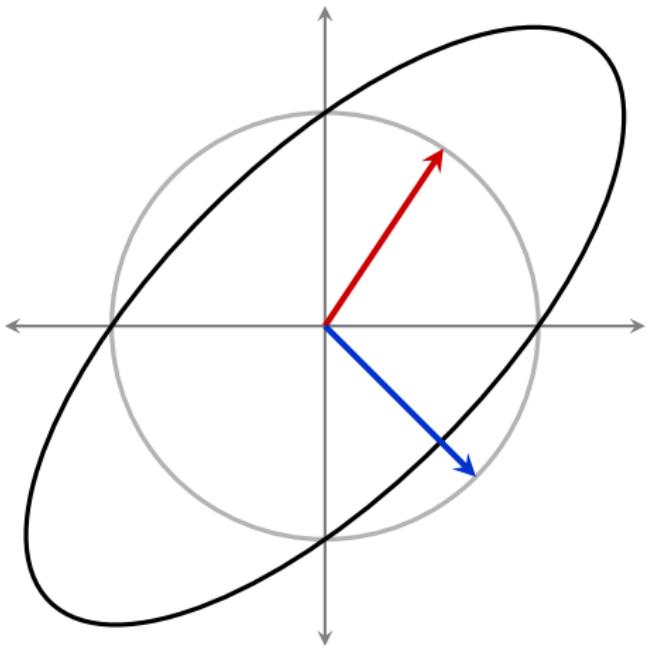
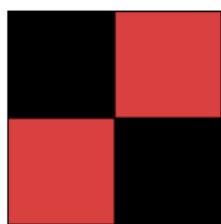
It's impossible to make a design perform better than independent assignment for all potential outcomes.



Details

Corollary

It's impossible to make a design perform better than independent assignment for all potential outcomes.



Details

Important takeaway

The design of experiments involves an **inescapable** trade-off between best-case and worst-case performance.

The only way to improve precision over independent assignment is by making the design less robust.

Misunderstandings between experimentalists and observationalists about causal inference

Kosuke Imai,

Princeton University, USA

Gary King

Harvard University, Cambridge, USA

and Elizabeth A. Stuart

Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

490 *K. Imai, G. King and E. A. Stuart*

founders are known and so can be adjusted for exactly, even if all covariates are not available. Thus, except in very small samples, **blocking on pretreatment variables followed by random treatment assignment cannot be worse than randomization alone**. Blocking on variables related to the outcome is of course more effective in increasing statistical efficiency than blocking on irrelevant variables, and so it pays to choose the variables to block carefully. But choosing not to block on a relevant pretreatment variable before randomization, that is feasible to use, is not justified.

5.2. Assumptions

Experimentalists and observationalists often make assumptions about unobserved processes on the basis of prior evidence or theory. At worst, when the question is judged to be sufficiently

Misunderstandings between experimentalists and observationalists about causal inference

Kosuke Imai,

Princeton University, USA

Gary King

Harvard University, Cambridge, USA

and Elizabeth A. Stuart

Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

490 *K. Imai, G. King and E. A. Stuart*

founders are known and so can be adjusted for exactly, even if all covariates are not available. Thus, except in very small samples, blocking on pretreatment variables followed by random treatment assignment ~~cannot~~ be worse than randomization alone. Blocking on variables related to the outcome is of ~~can~~ more effective in increasing statistical efficiency than blocking on irrelevant variables, and so it pays to choose the variables to block carefully. But choosing not to block on a relevant pretreatment variable before randomization, that is feasible to use, is not justified.

5.2. Assumptions

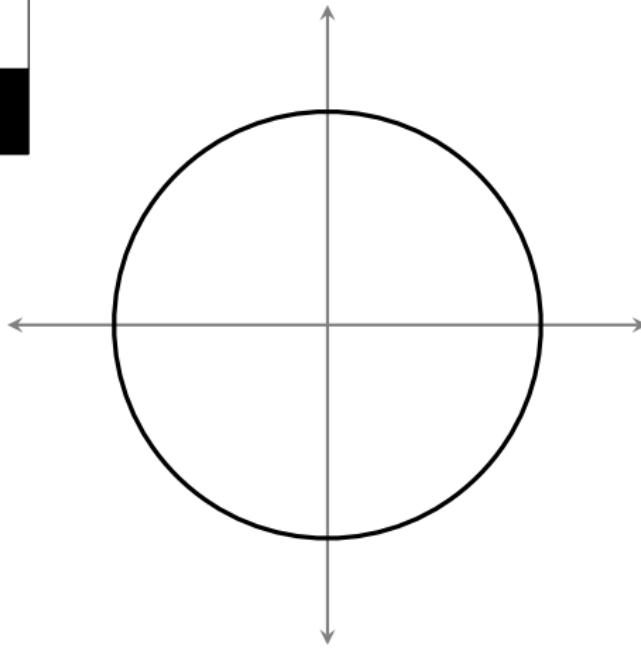
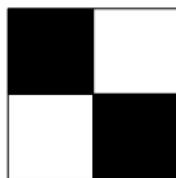
Experimentalists and observationalists often make assumptions about unobserved processes on the basis of prior evidence or theory. At worst, when the question is judged to be sufficiently

Theorem

As the design becomes more deterministic, the ellipse becomes more squeezed.

Theorem

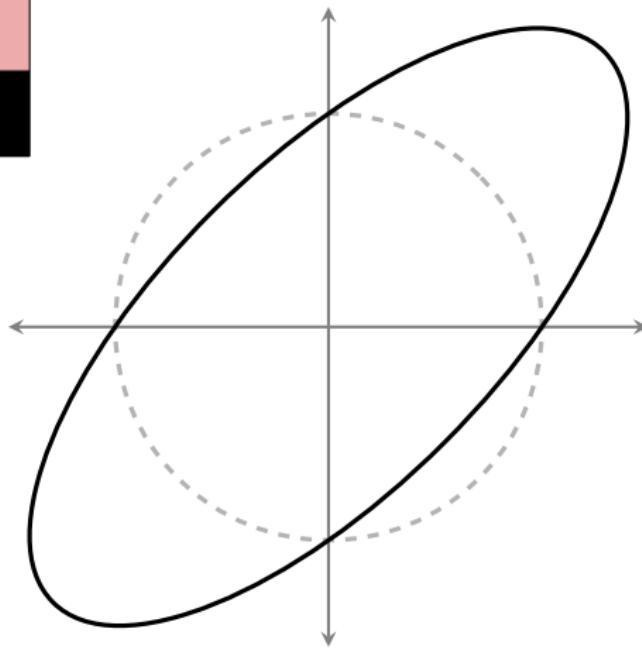
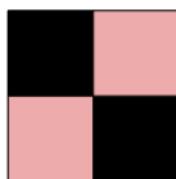
As the design becomes more deterministic, the ellipse becomes more squeezed.



Details

Theorem

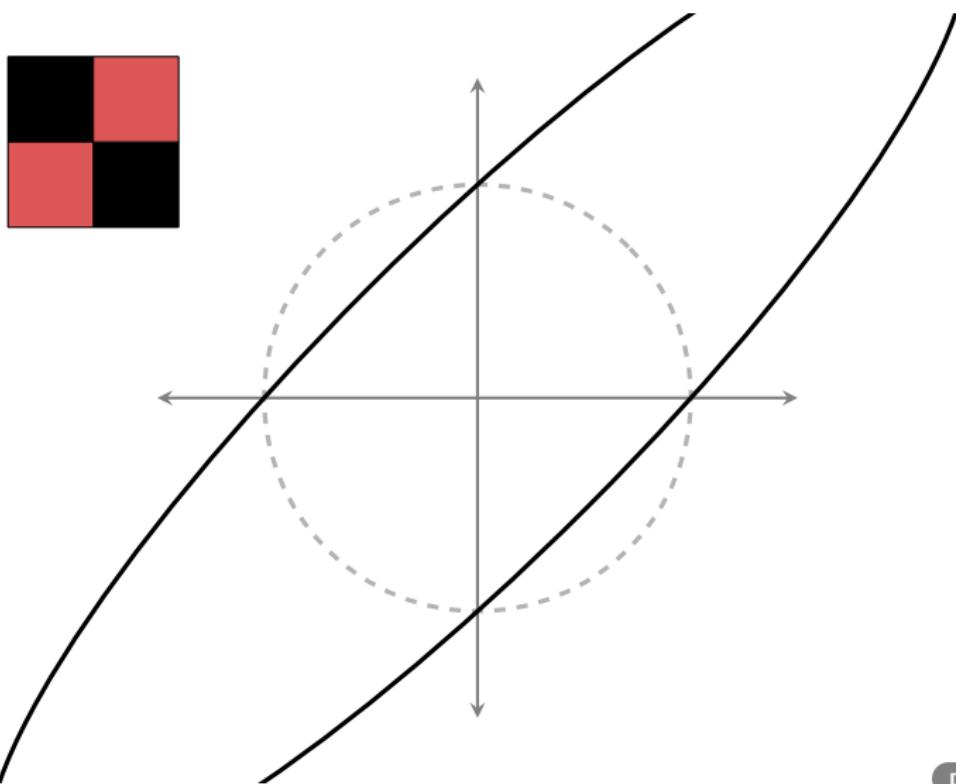
As the design becomes more deterministic, the ellipse becomes more squeezed.



Details

Theorem

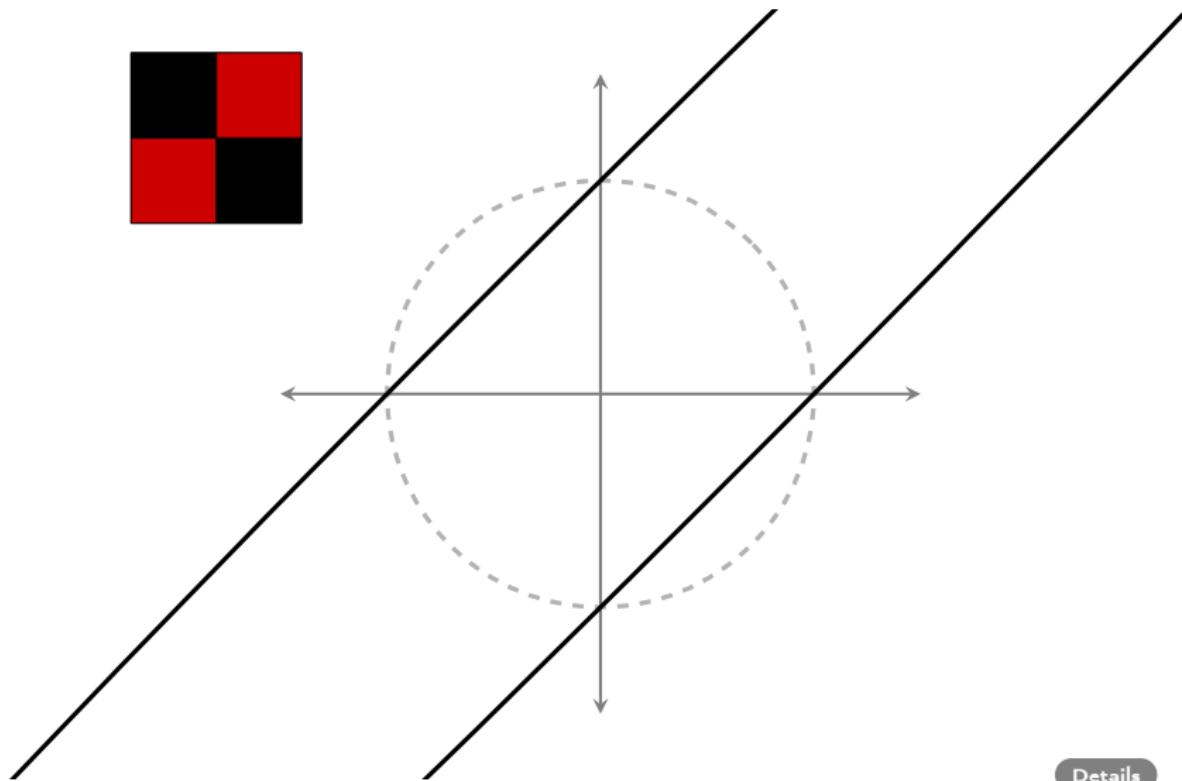
As the design becomes more deterministic, the ellipse becomes more squeezed.



Details

Theorem

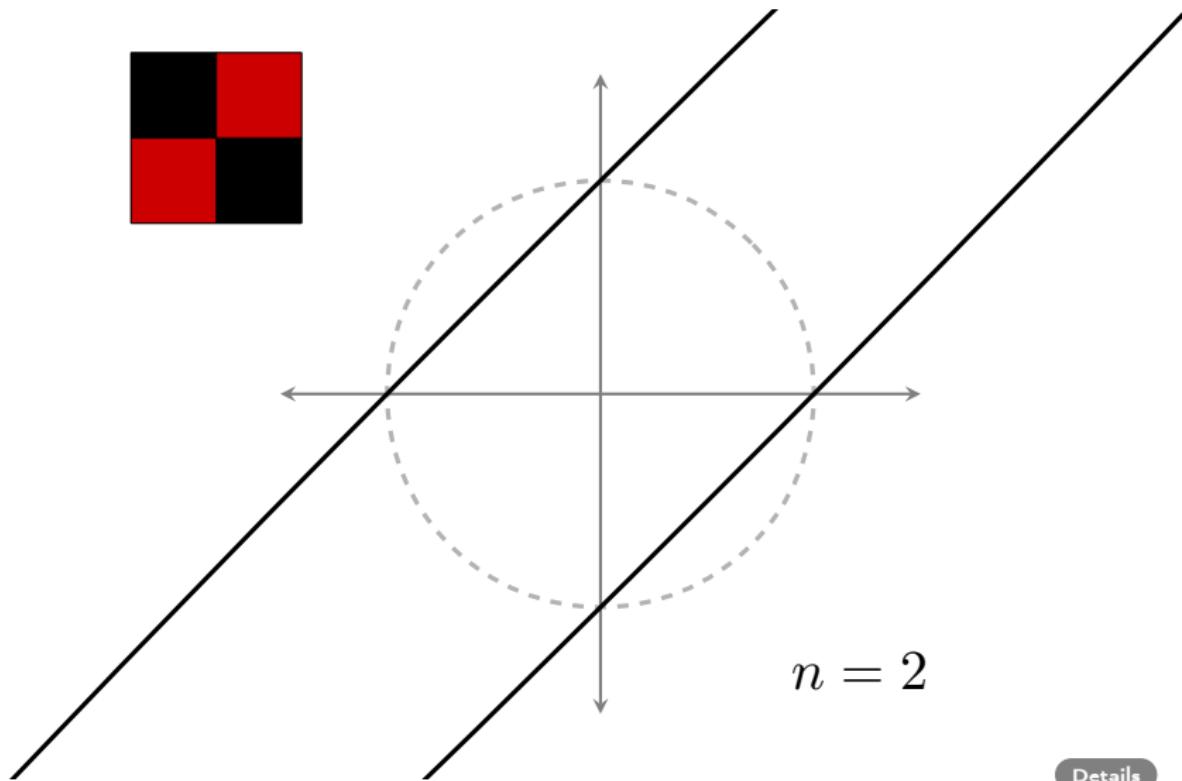
As the design becomes more deterministic, the ellipse becomes more squeezed.



Details

Theorem

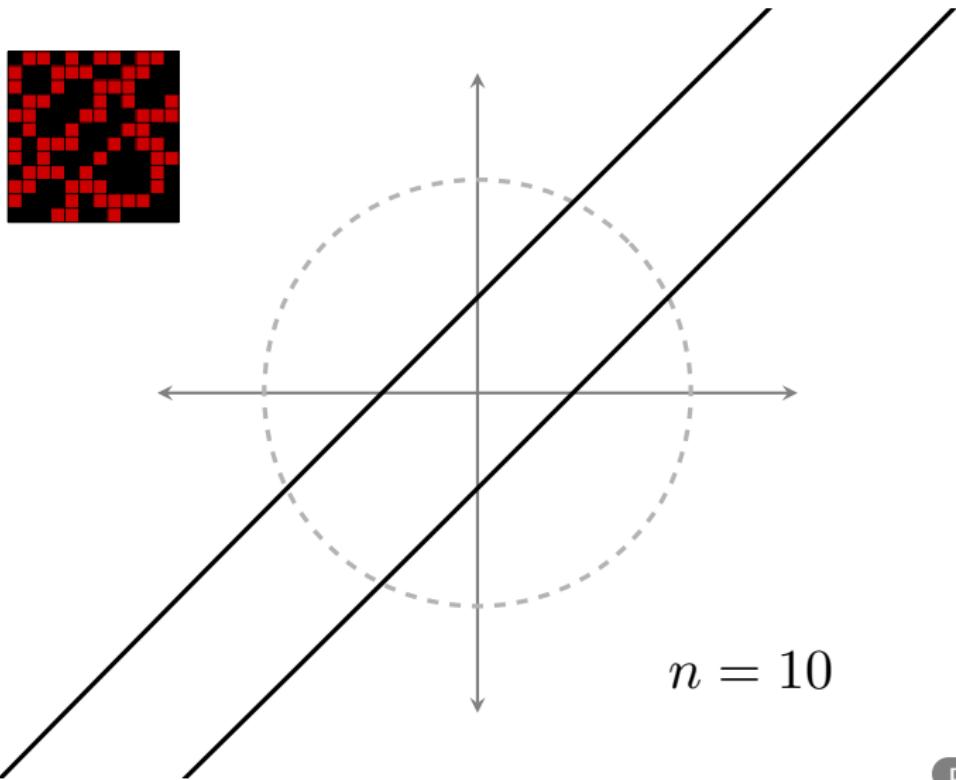
As the design becomes more deterministic, the ellipse becomes more squeezed.



Details

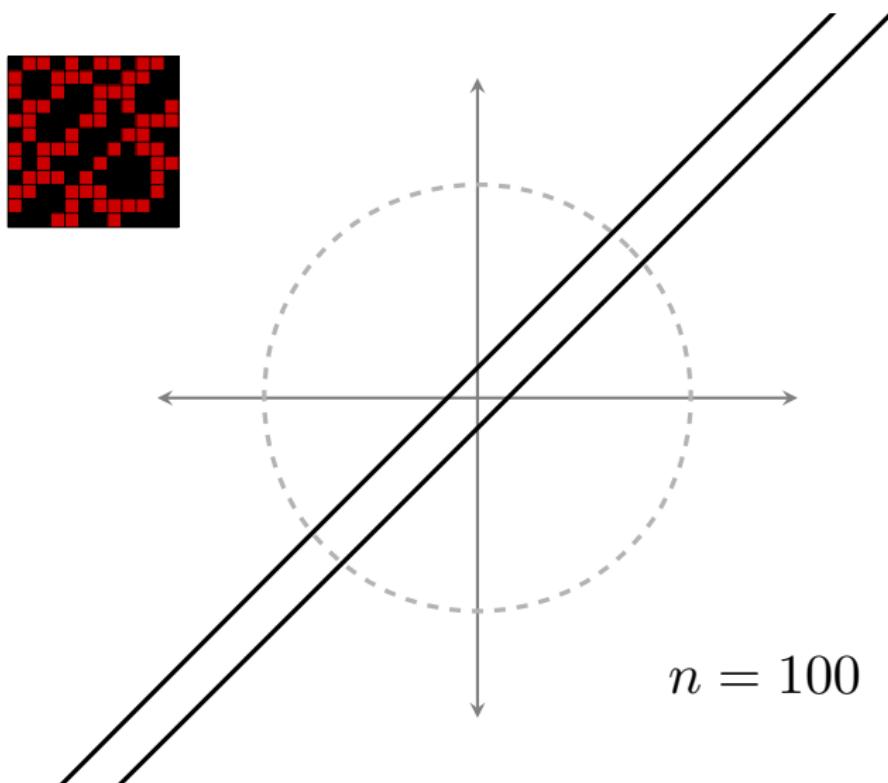
Theorem

As the design becomes more deterministic, the ellipse becomes more squeezed.



Theorem

As the design becomes more deterministic, the ellipse becomes more squeezed.



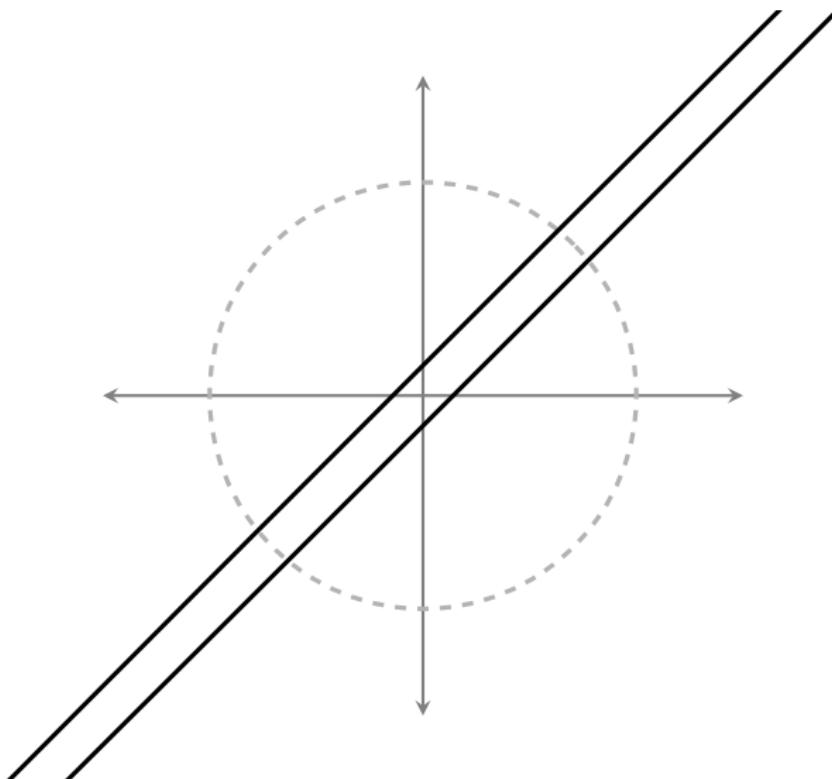
Details

Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.

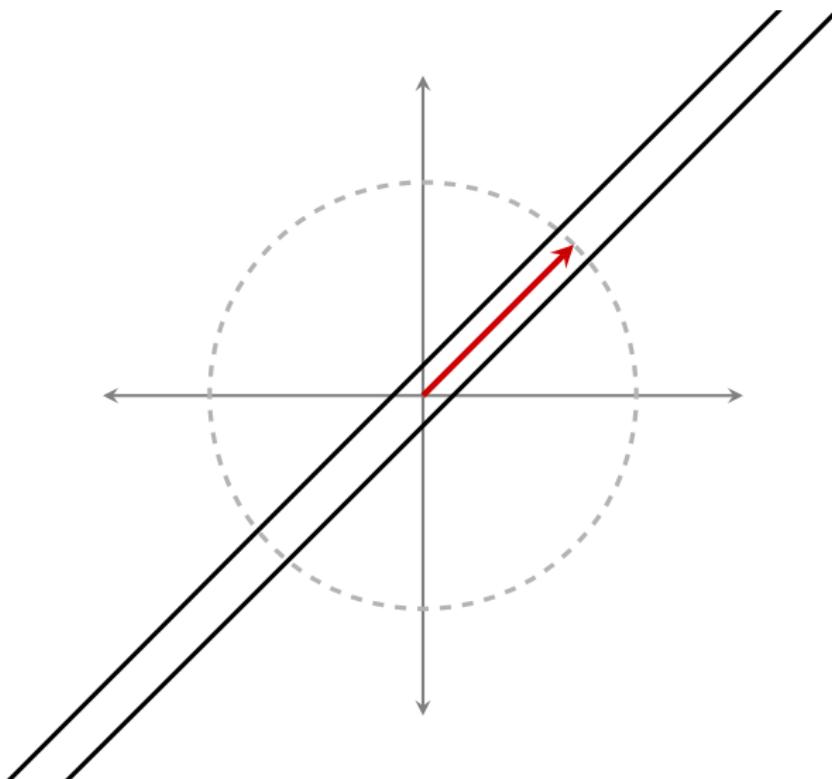
Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.



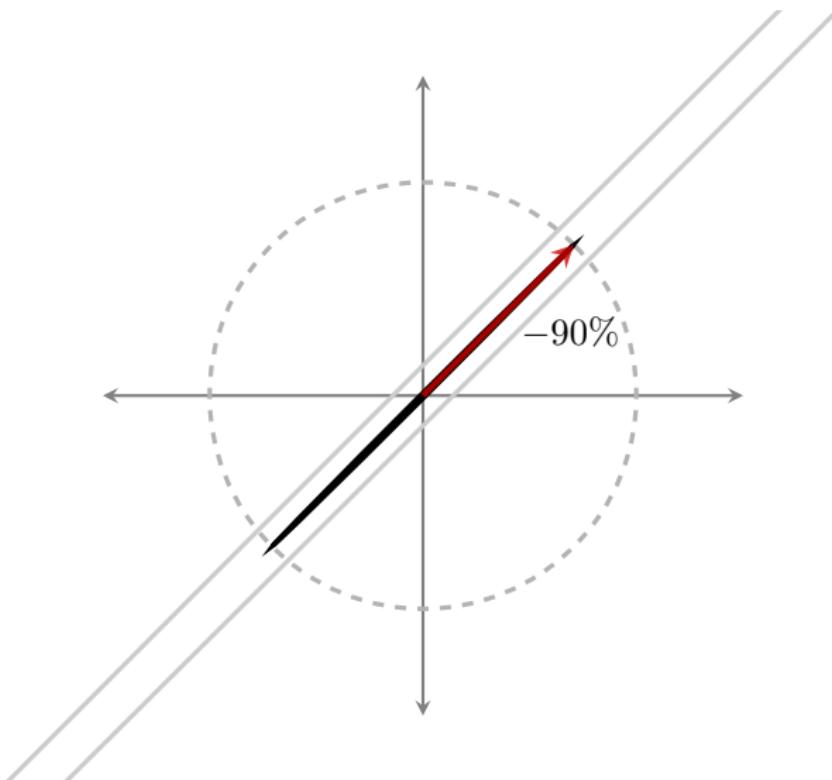
Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.



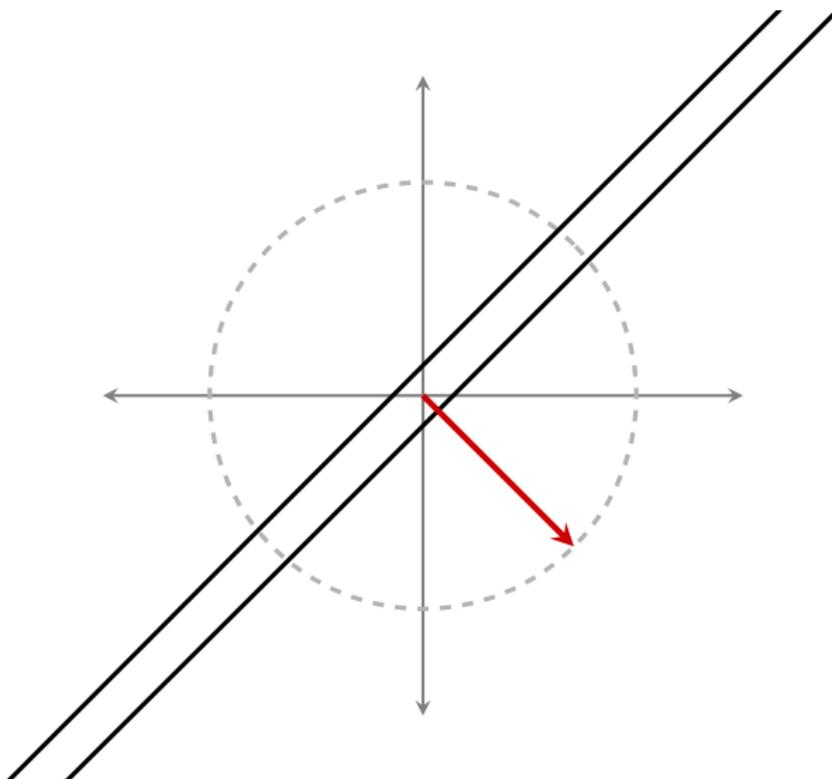
Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.



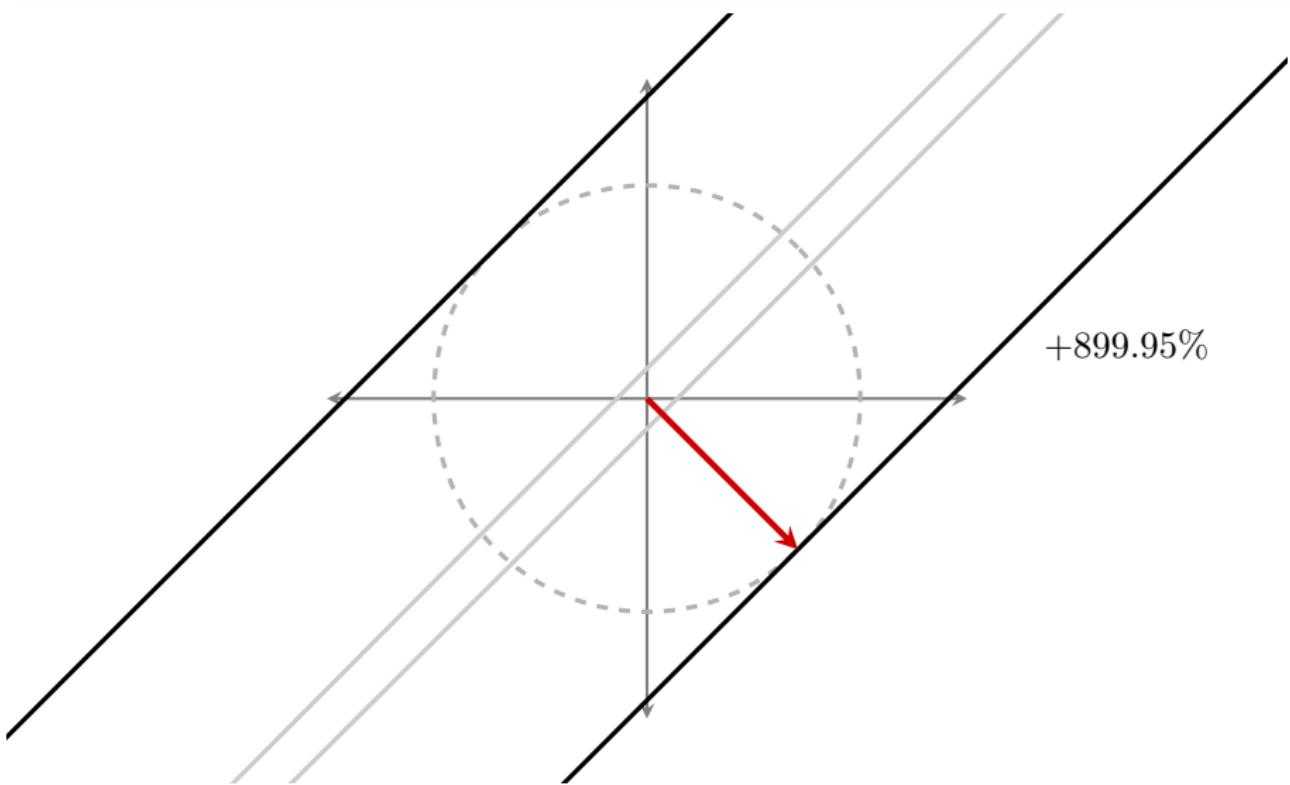
Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.



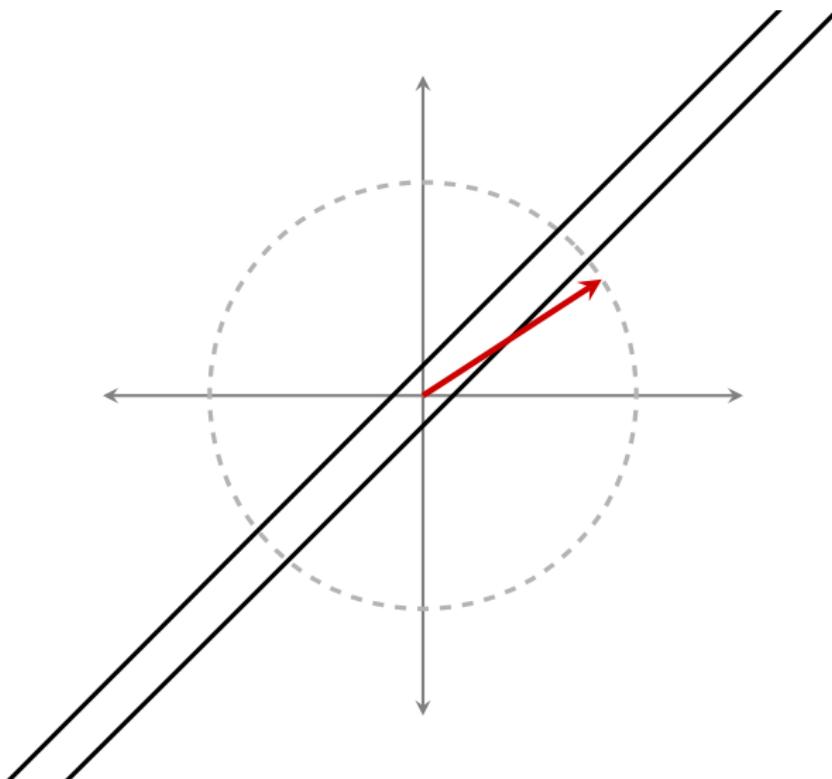
Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.



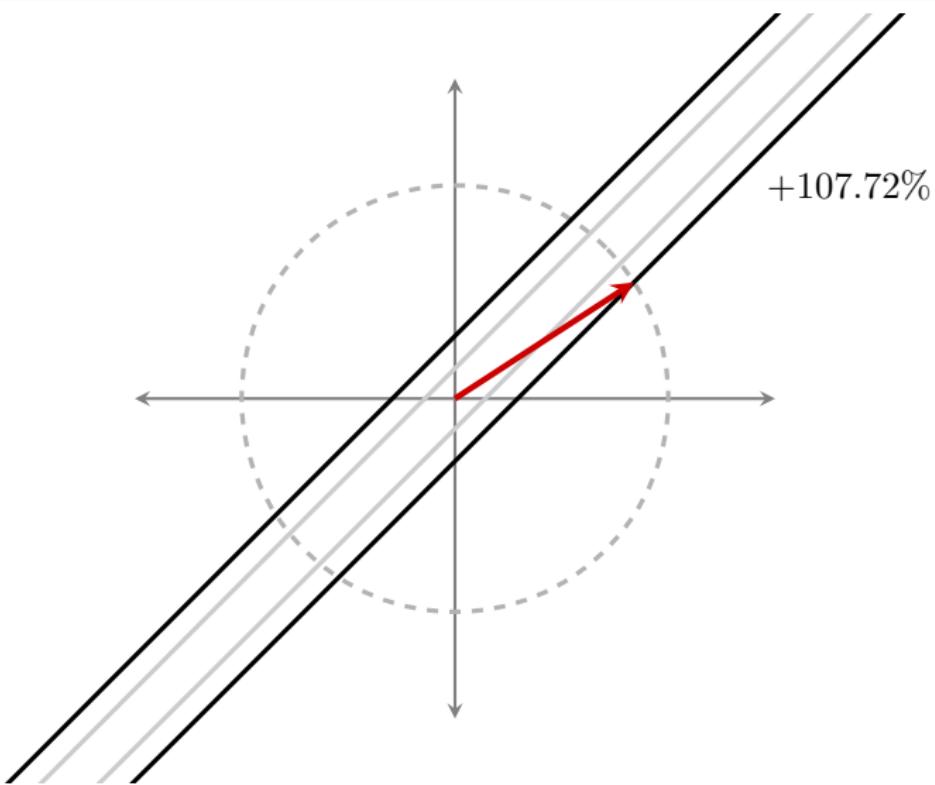
Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.



Corollary

If a design is close to deterministic, it places an enormous bet on getting the direction of the potential outcomes right.

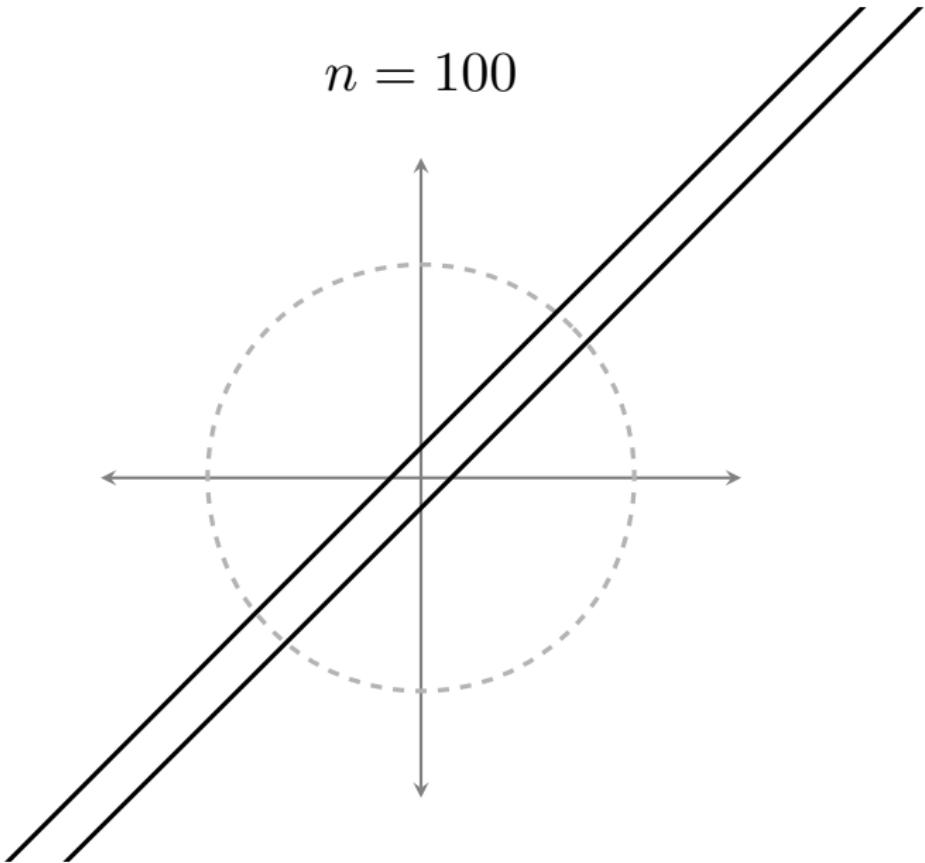


Theorem

If a design squeezes the ellipse too much, the estimation error will generally not decrease as the sample grows larger.

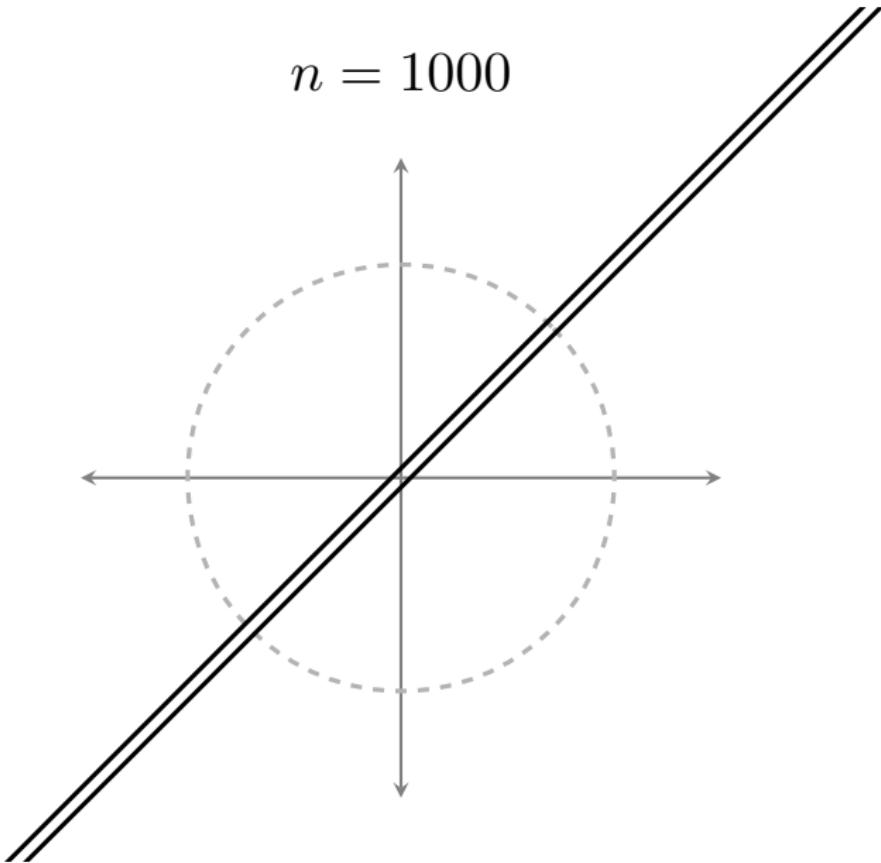
That is, the estimator will not be consistent if the design is close to deterministic.

$$n = 100$$



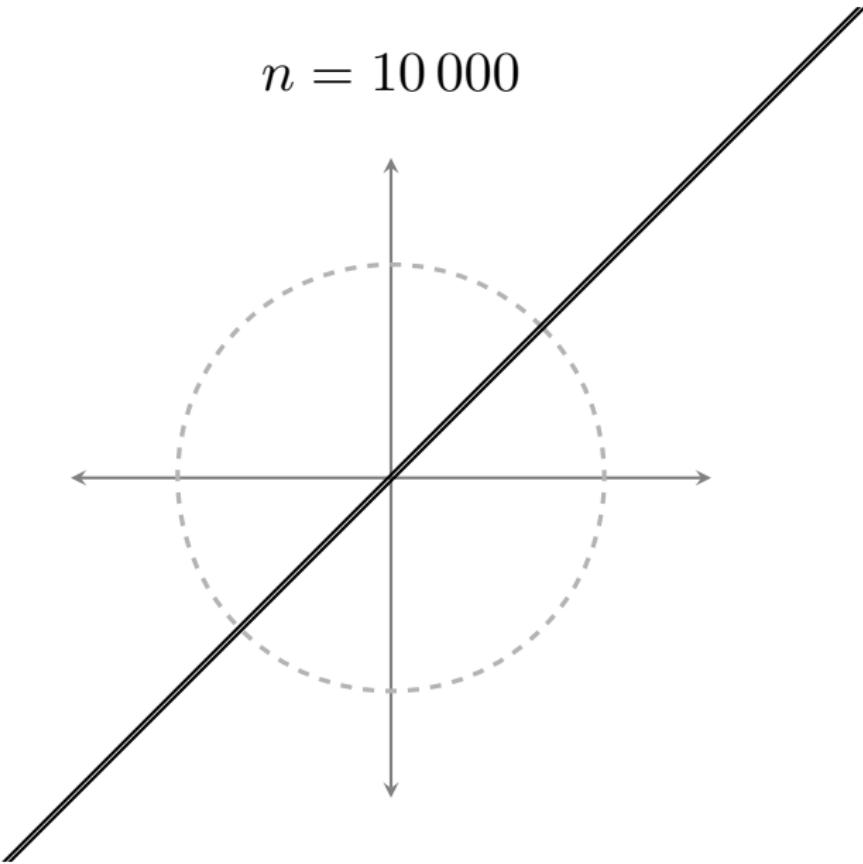
Note: Low-dimensional view.

$$n = 1000$$



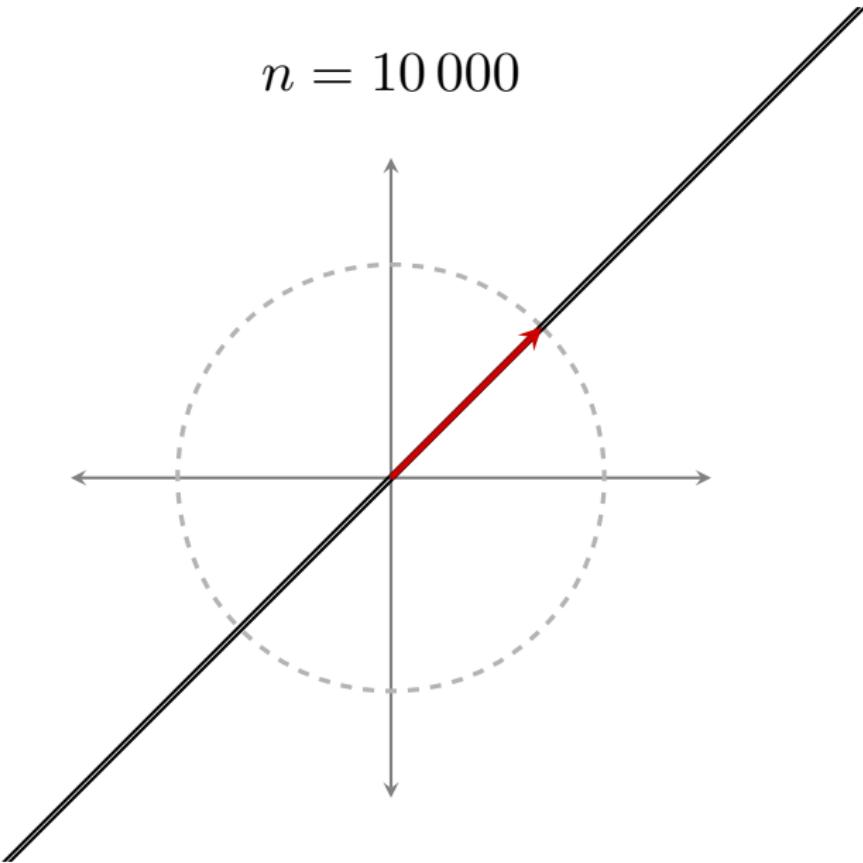
Note: Low-dimensional view.

$$n = 10\,000$$



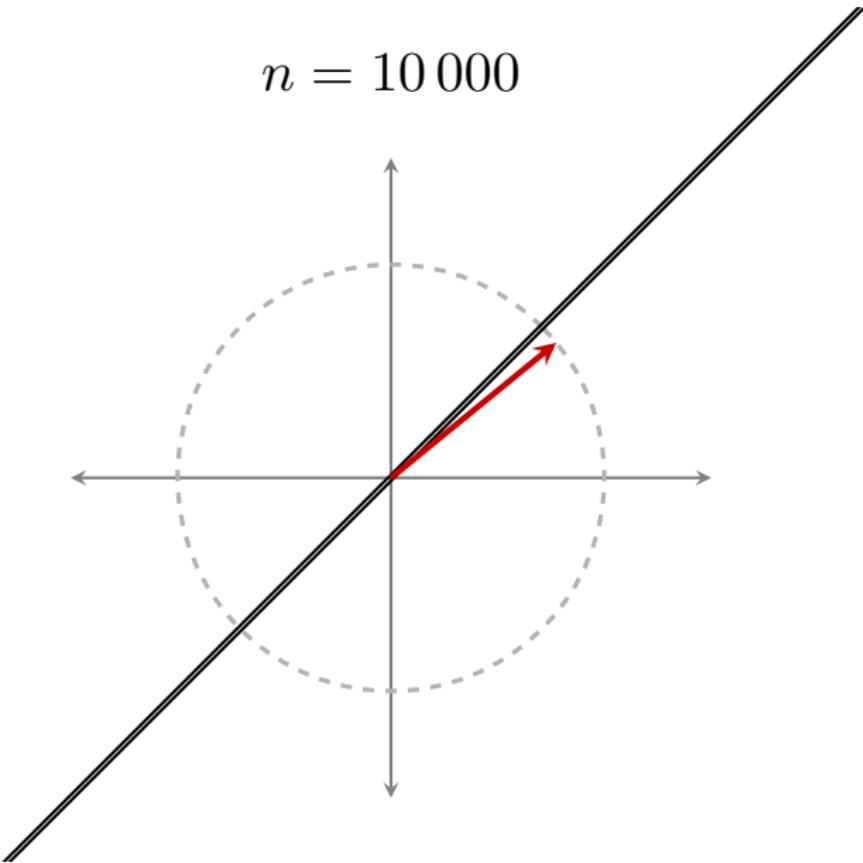
Note: Low-dimensional view.

$$n = 10\,000$$



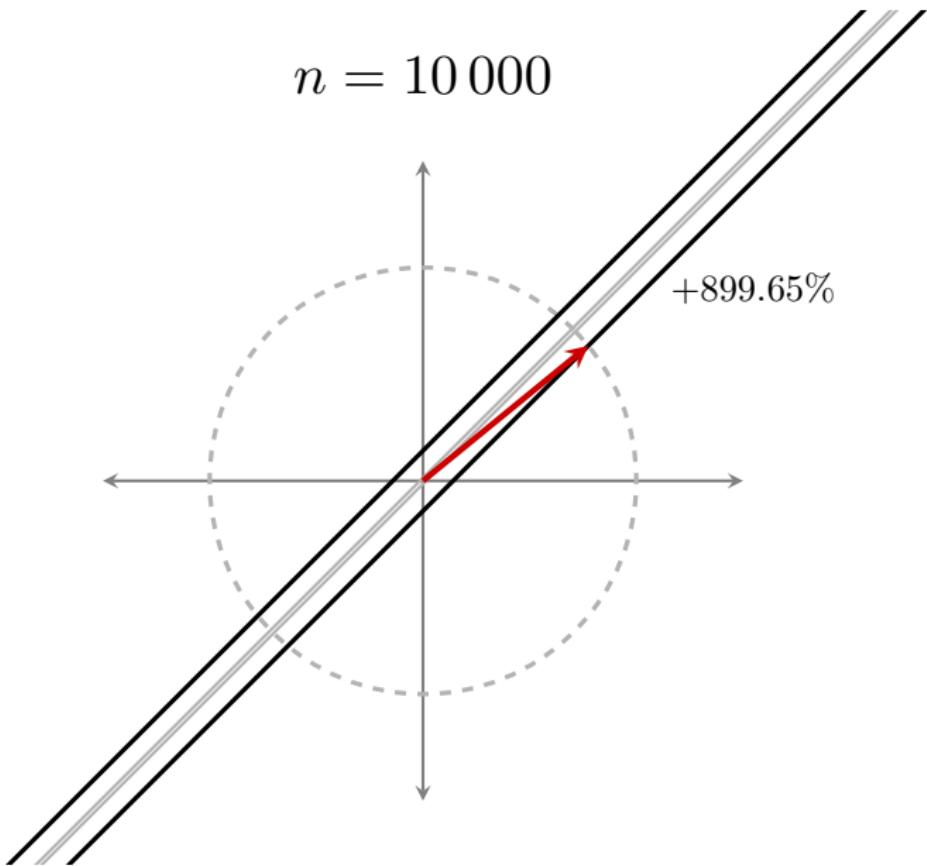
Note: Low-dimensional view.

$$n = 10\,000$$

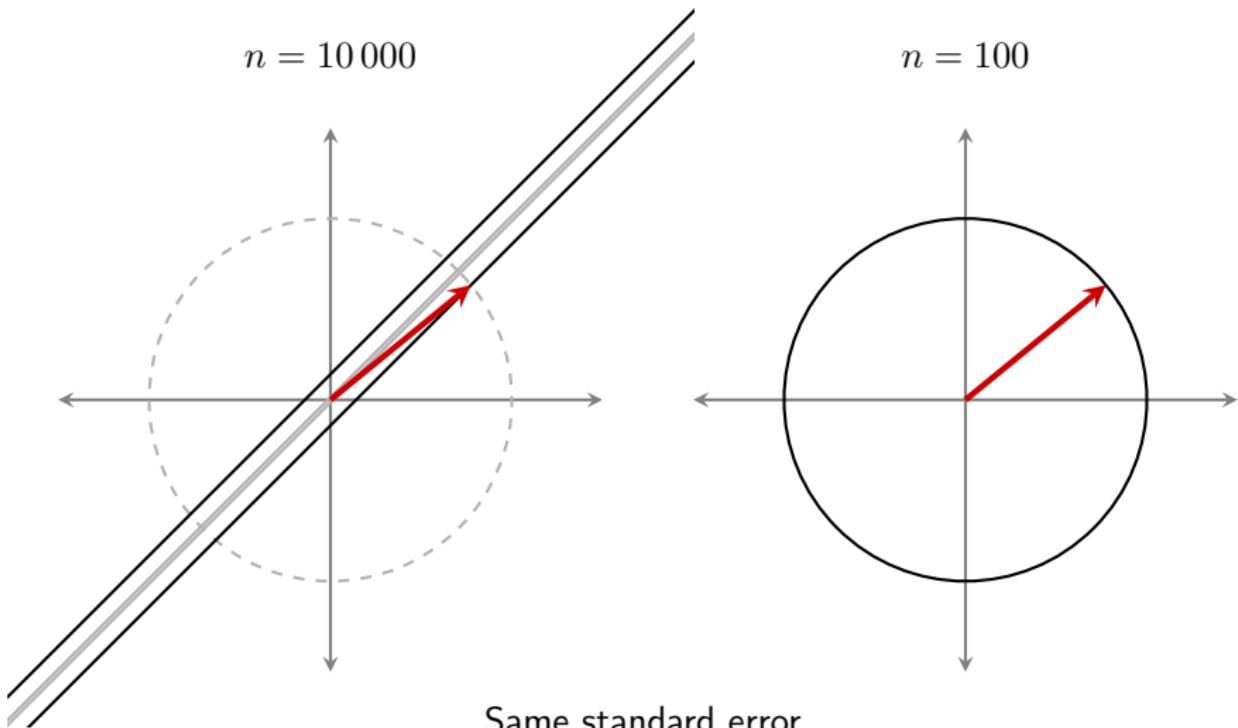


Note: Low-dimensional view.

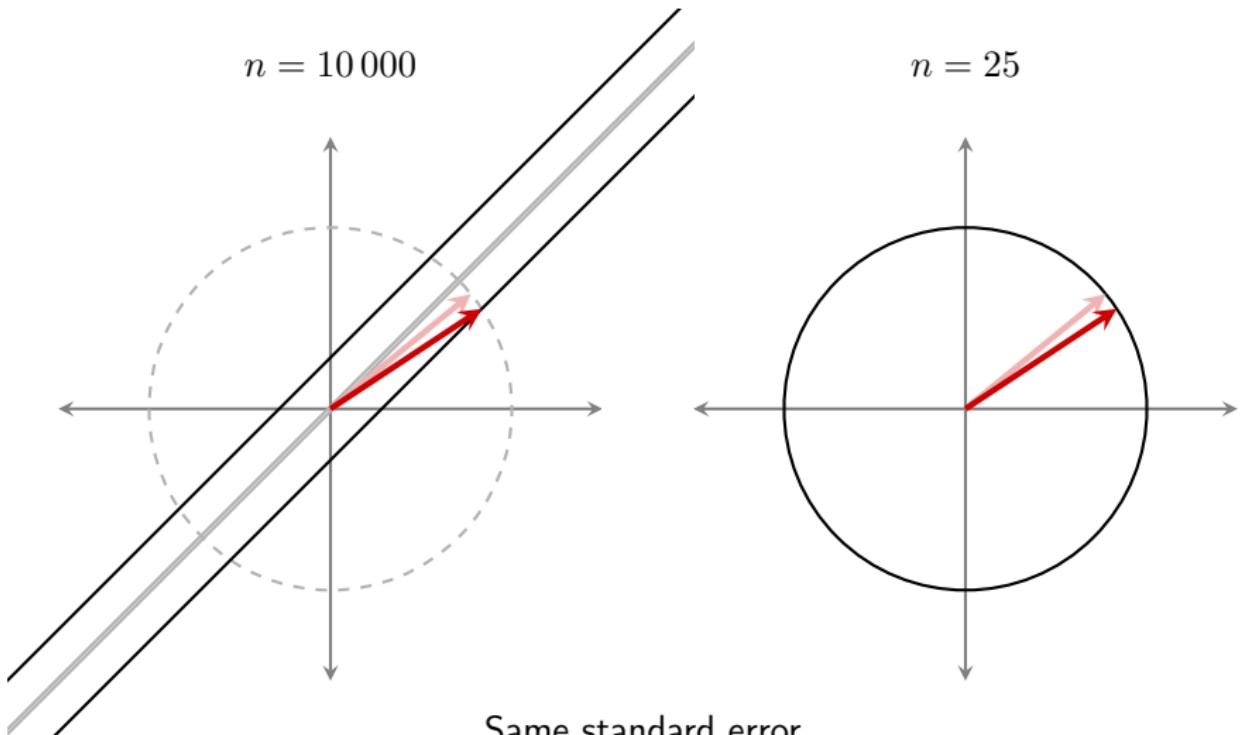
$n = 10\,000$



Note: Low-dimensional view.



Note: Low-dimensional view.



Note: Low-dimensional view.

Important takeaway

We randomize because we don't know the direction of the potential outcomes, so we want hedge our bet.

Deterministic assignment performs well **only** if we get the direction right.

Randomization performs reasonably well even if we get it wrong.

Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead

Maximilian Kasy

*Department of Economics, Harvard University, 1805 Cambridge Street,
Cambridge, MA 02138, USA
e-mail: maximilankasy@fas.harvard.edu (corresponding author)*

Suppose that an experimenter has collected a sample as well as baseline information about the units in the sample. How should she allocate treatments to the units in this sample? We argue that the answer does not involve randomization if we think of experimental design as a statistical decision problem. If, for instance, the experimenter is interested in estimating the average treatment effect and evaluates an estimate in terms of the squared error, then she should minimize the expected mean squared error (MSE) through choice of a treatment assignment. We provide explicit expressions for the expected MSE that lead to easily implementable procedures for experimental design.

1 Introduction

Experiments, and in particular randomized experiments, are the conceptual reference point that gives empirical content to the notion of causality. In recent years, actual randomized experiments have become increasingly popular elements of the methodological toolbox in a wide range of social science disciplines. Examples from the recent political science literature abound. Blattman

Why Experimenters Might ~~Not~~ Always Want to Randomize, ~~and What They Could Do Instead~~

Maximilian Kasy

*Department of Economics, Harvard University, 1805 Cambridge Street,
Cambridge, MA 02138, USA
e-mail: maximilankasy@fas.harvard.edu (corresponding author)*

Suppose that an experimenter has collected a sample as well as baseline information about the units in the sample. How should she allocate treatments to the units in this sample? We argue that the answer does not involve randomization if we think of experimental design as a statistical decision problem. If, for instance, the experimenter is interested in estimating the average treatment effect and evaluates an estimate in terms of the squared error, then she should minimize the expected mean squared error (MSE) through choice of a treatment assignment. We provide explicit expressions for the expected MSE that lead to easily implementable procedures for experimental design.

1 Introduction

Experiments, and in particular randomized experiments, are the conceptual reference point that gives empirical content to the notion of causality. In recent years, actual randomized experiments have become increasingly popular elements of the methodological toolbox in a wide range of social science disciplines. Examples from the recent political science literature abound. Blattman

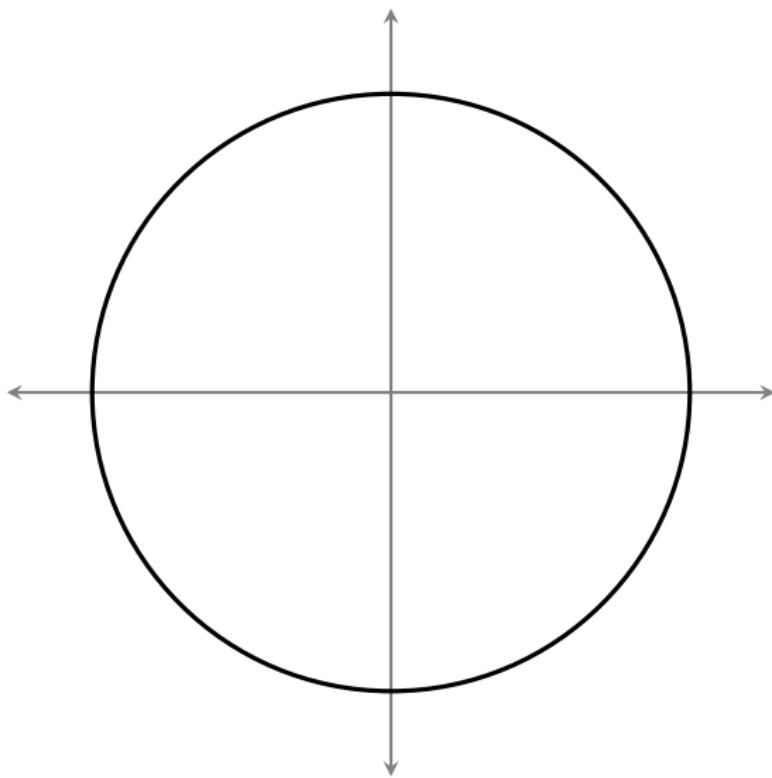
Today's talk

- Why should we randomize? ✓
- How should we randomize?
- How can we randomize in practice?

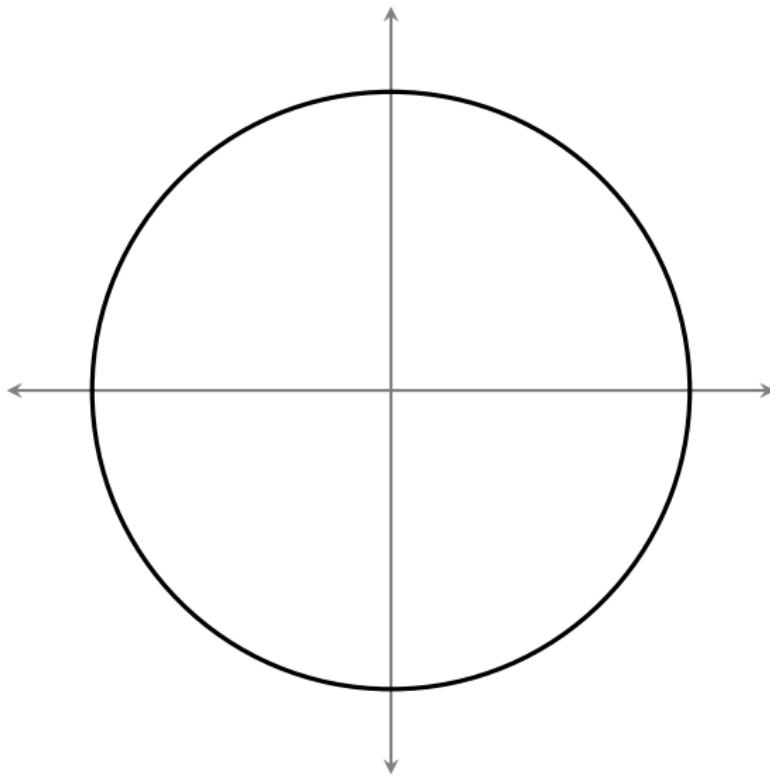
Today's talk

- Why should we randomize? ✓
- How should we randomize?
- How can we randomize in practice?

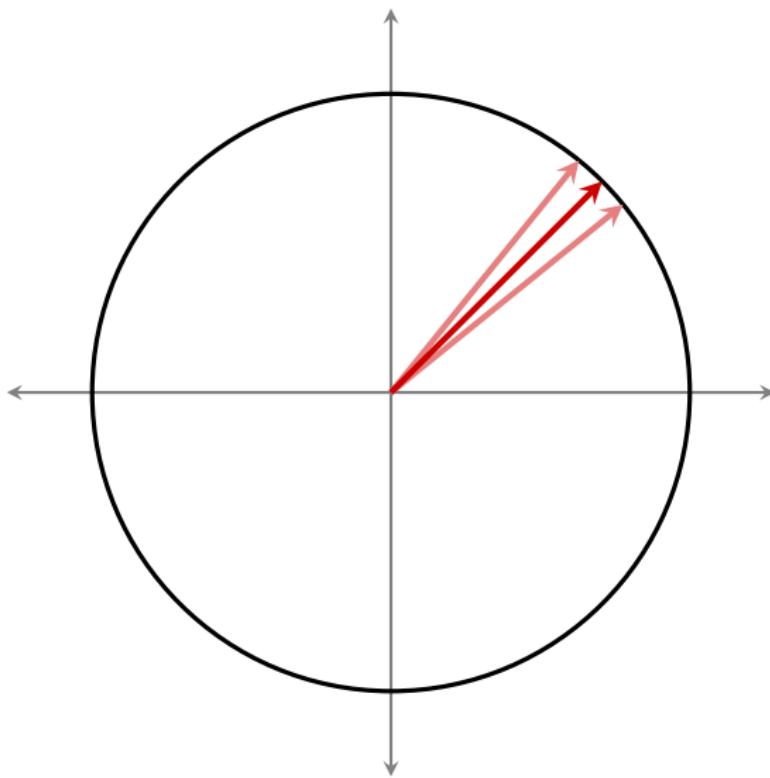
So, always independent assignment?



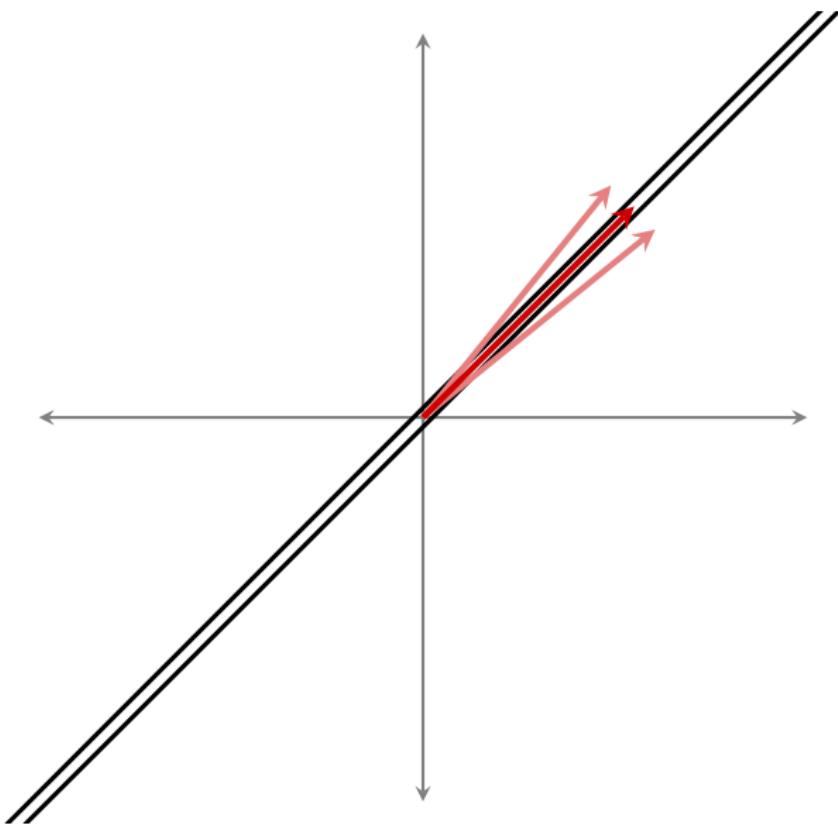
So, always independent assignment? No!



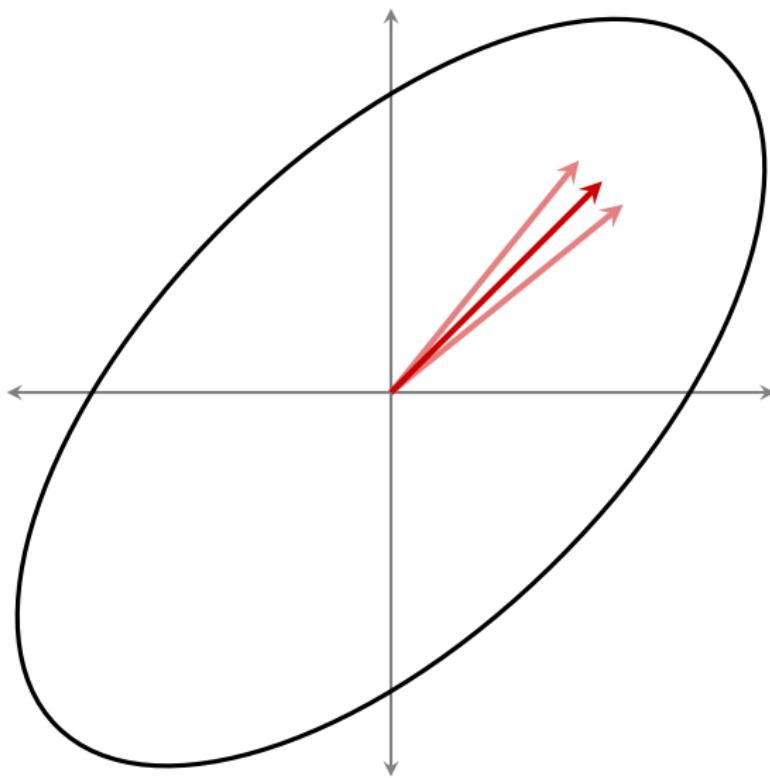
So, always independent assignment? No!



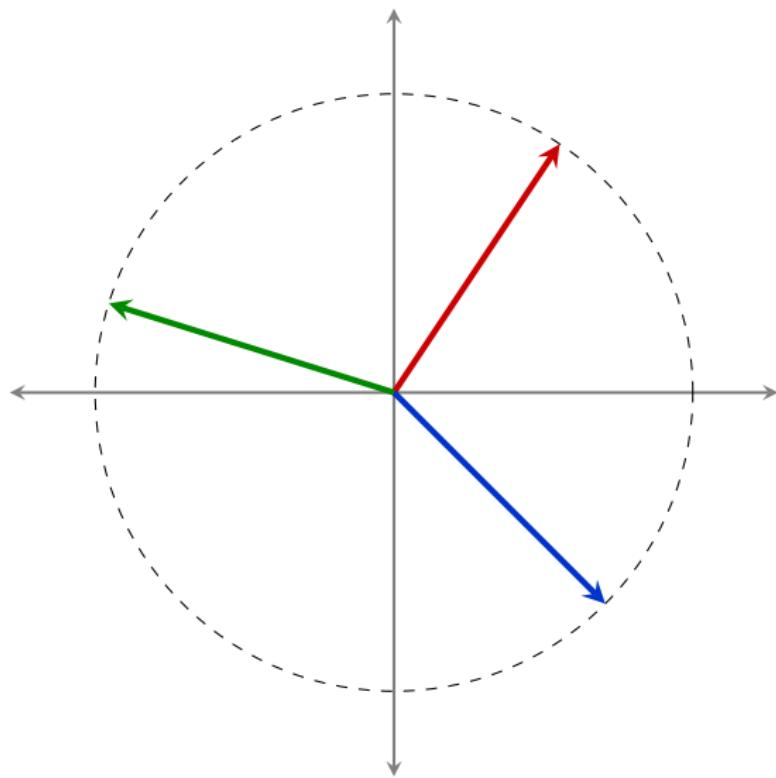
So, always independent assignment? No!



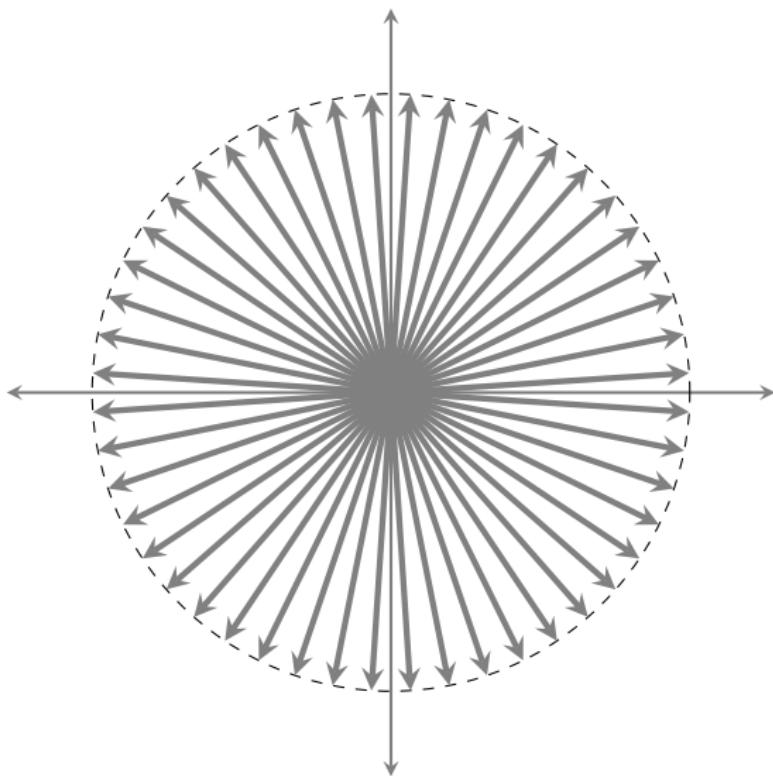
So, always independent assignment? No!



But we don't know the potential outcomes



But we don't know the potential outcomes

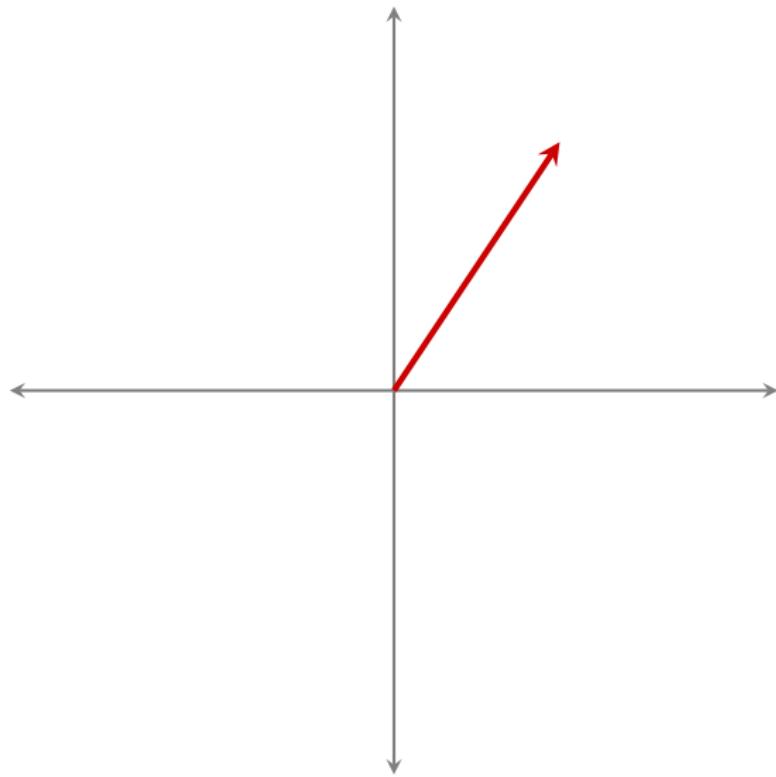


What is known before assignment?

We know baseline covariates X .

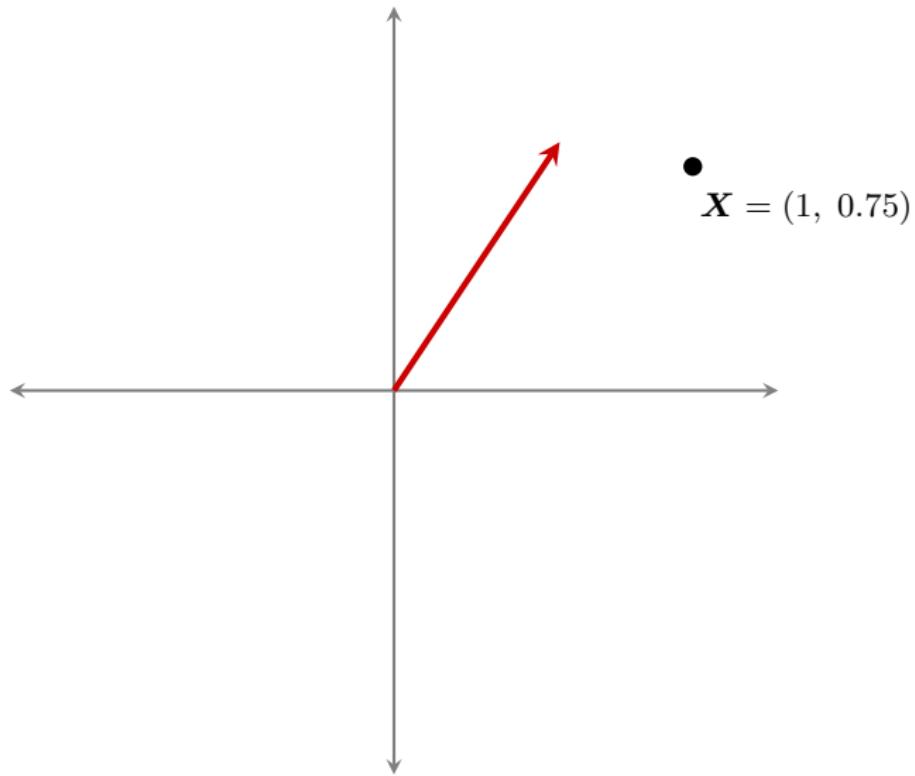
We can use the covariates to decompose
the potential outcomes into two **known** directions.

Decomposing the outcomes into two known directions



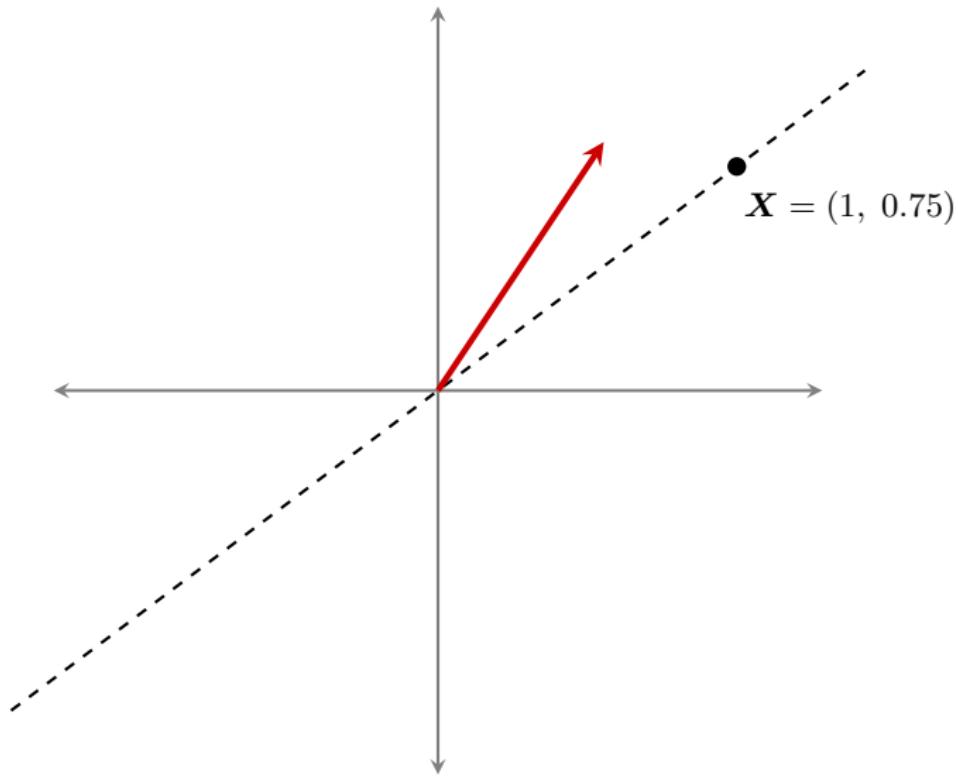
Details

Decomposing the outcomes into two known directions



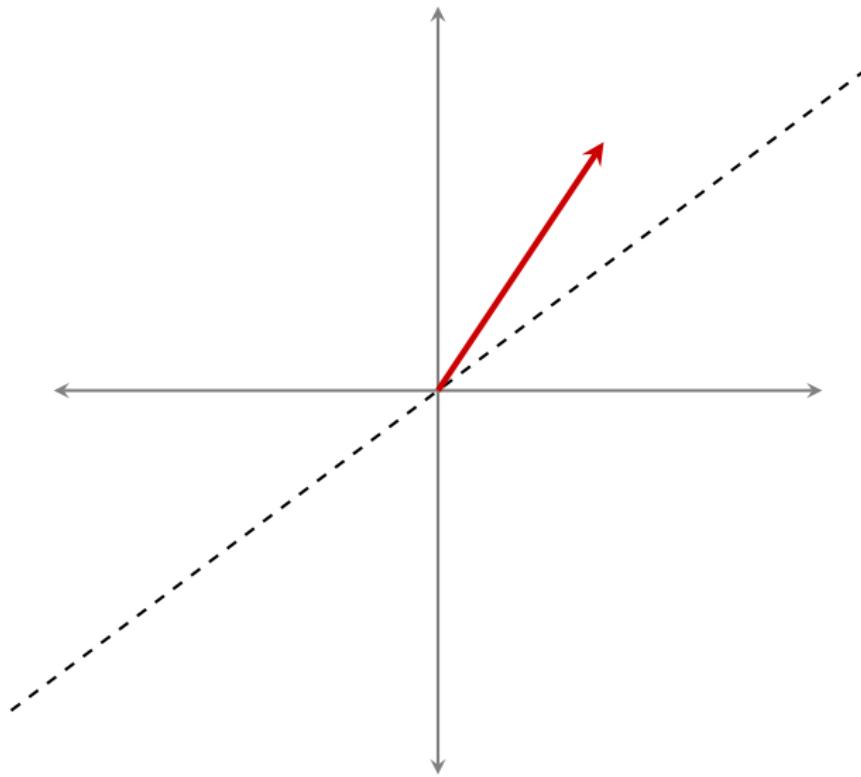
Details

Decomposing the outcomes into two known directions



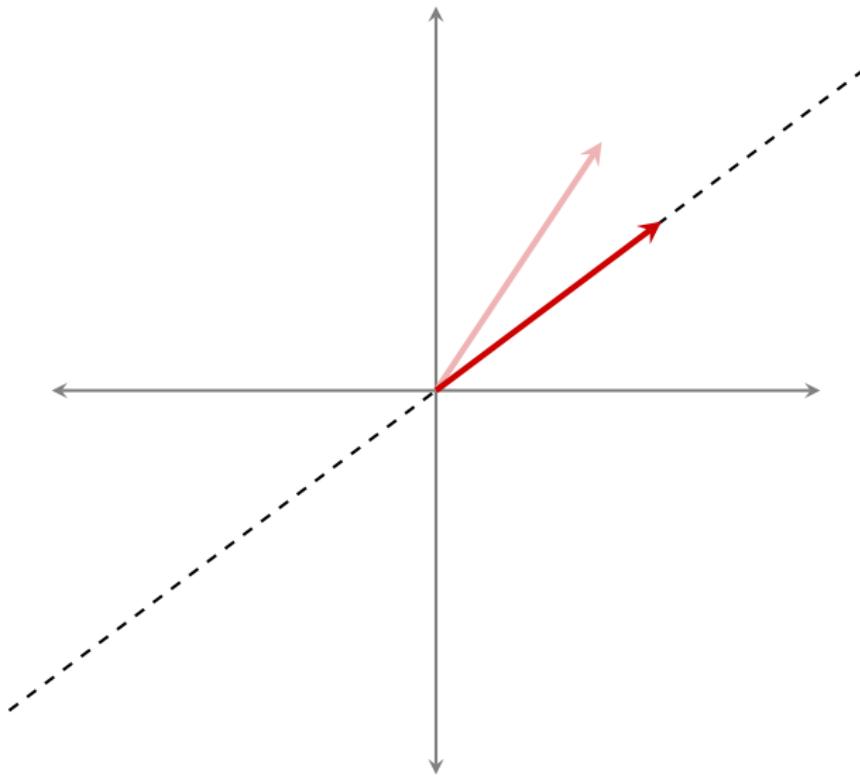
Details

Decomposing the outcomes into two known directions



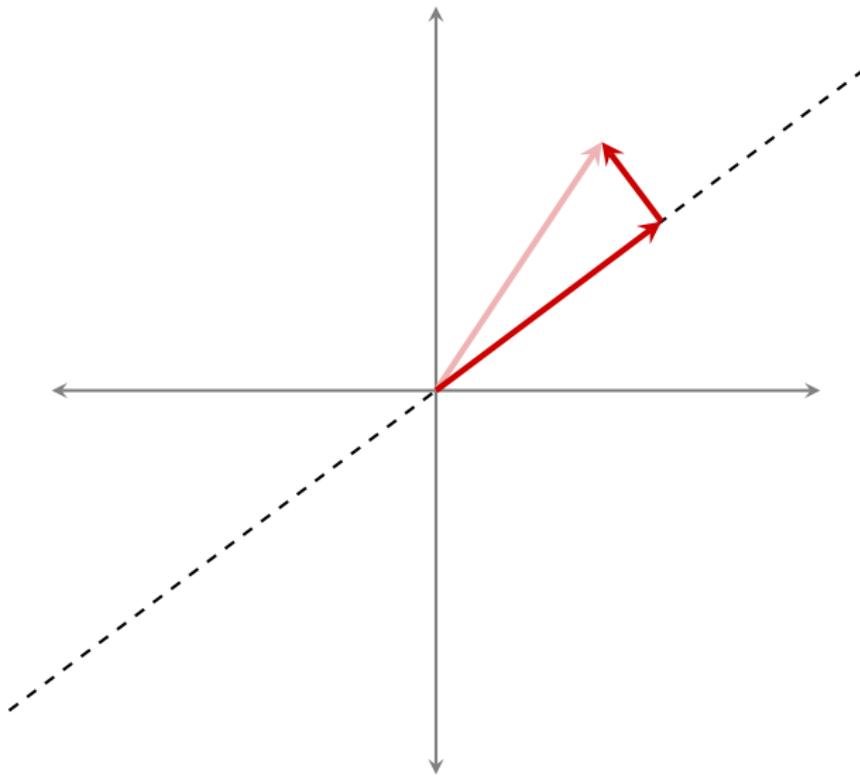
Details

Decomposing the outcomes into two known directions



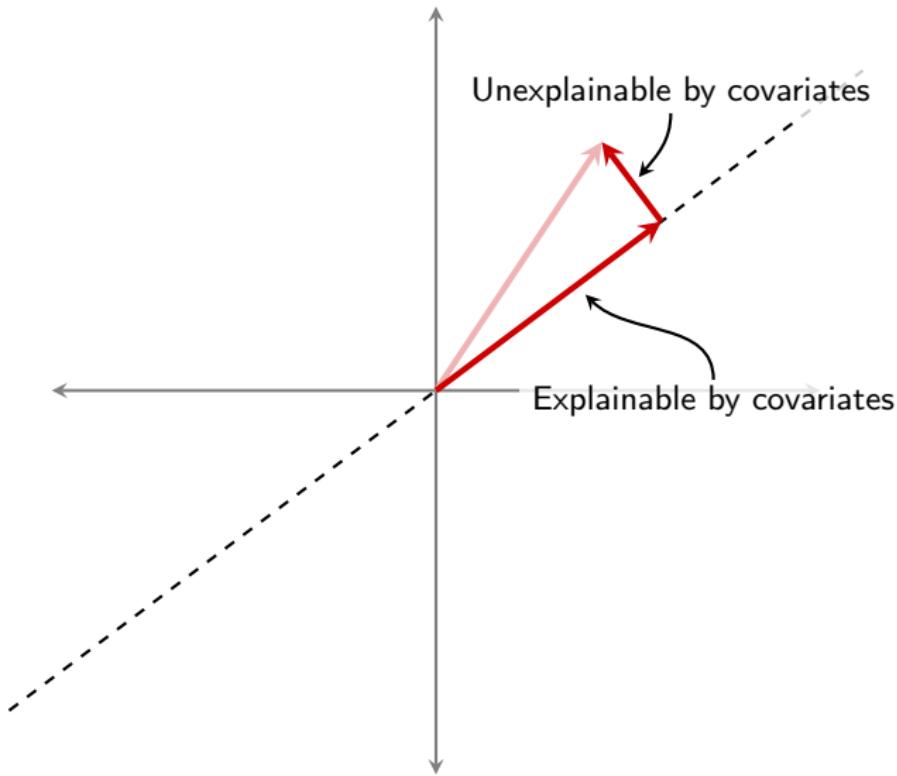
Details

Decomposing the outcomes into two known directions



Details

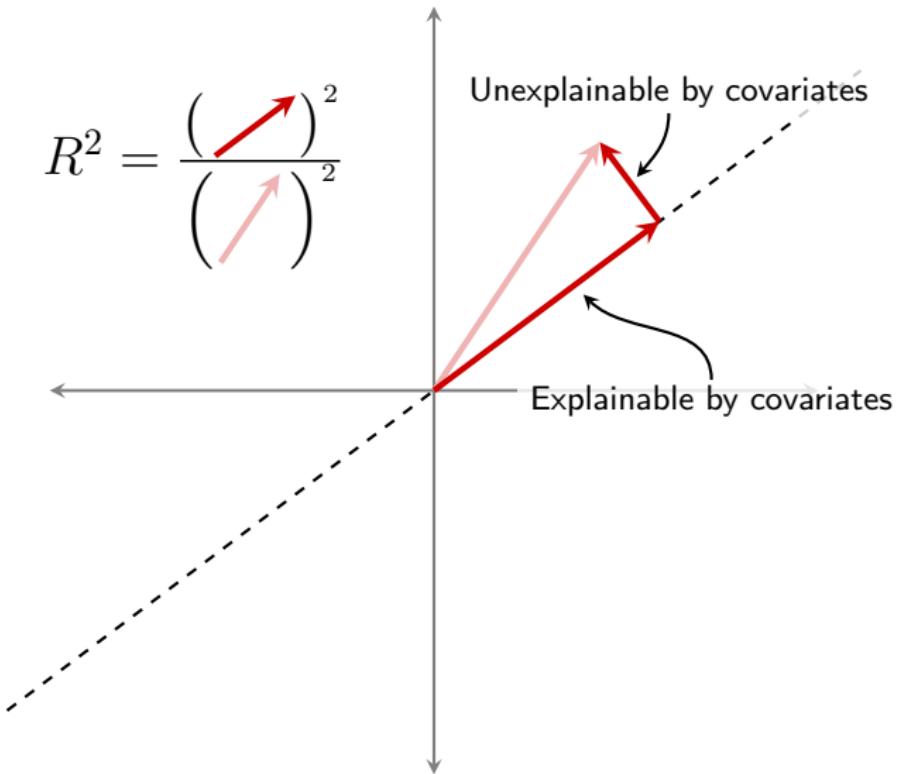
Decomposing the outcomes into two known directions



Details

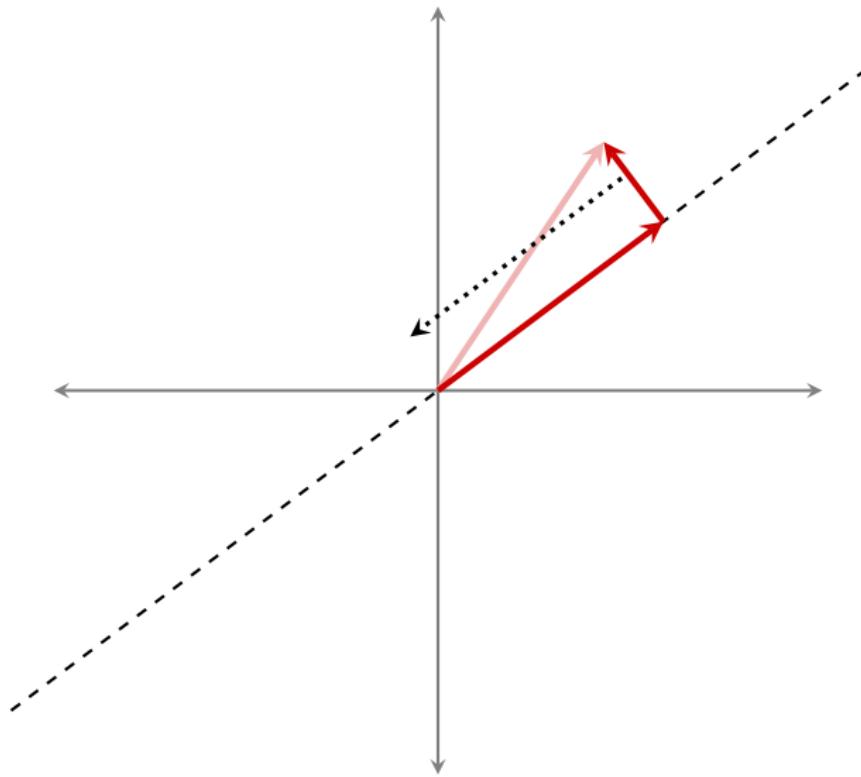
Decomposing the outcomes into two known directions

$$R^2 = \frac{(\text{Red Arrow})^2}{(\text{Total Length})^2}$$



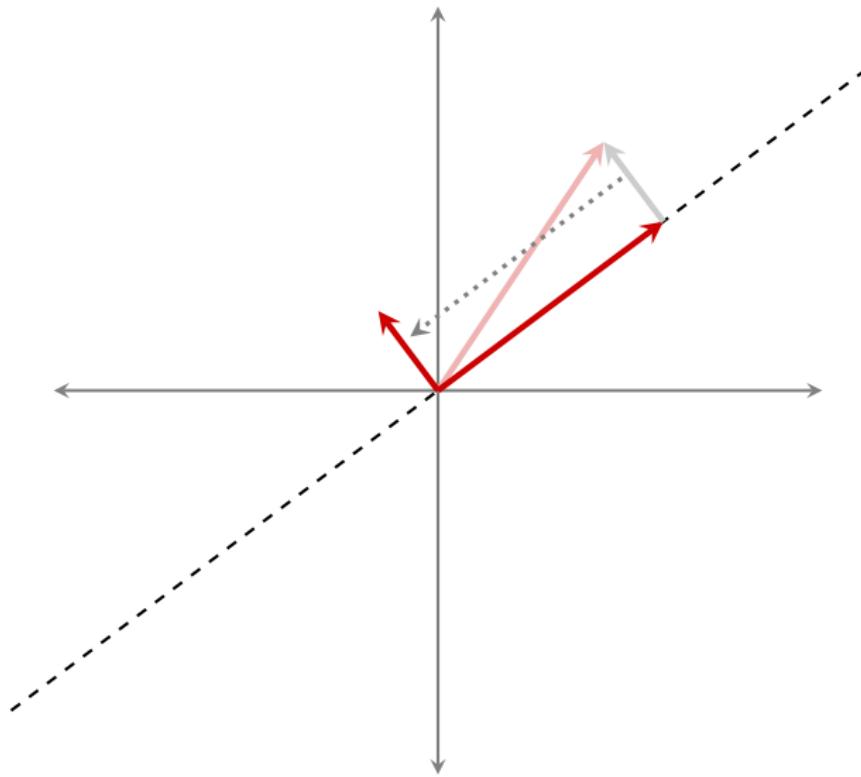
Details

Decomposing the outcomes into two known directions



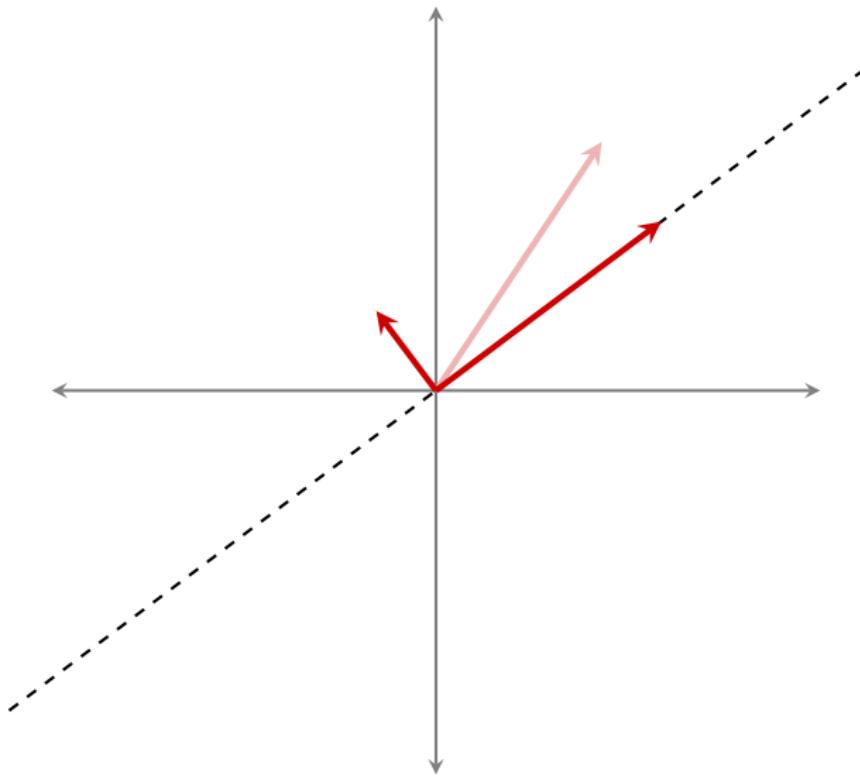
Details

Decomposing the outcomes into two known directions



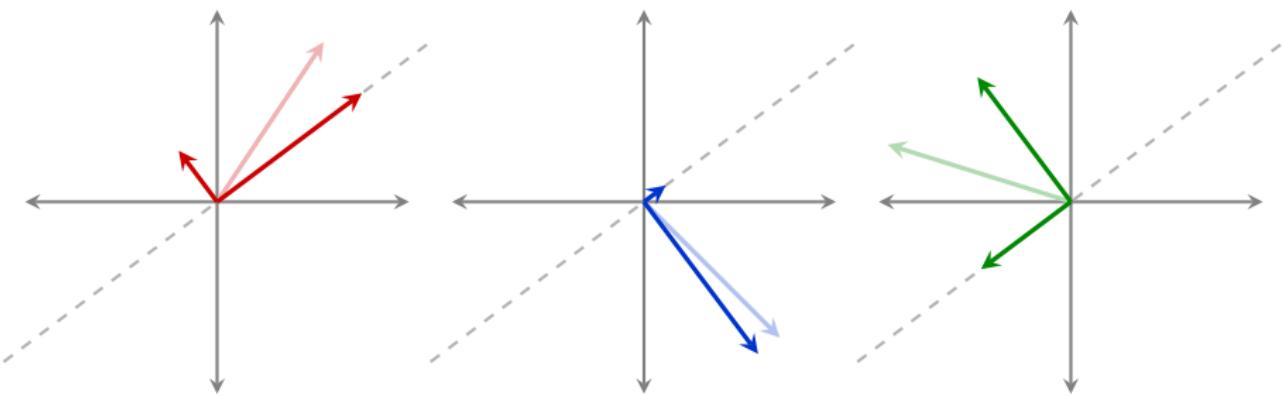
Details

Decomposing the outcomes into two known directions

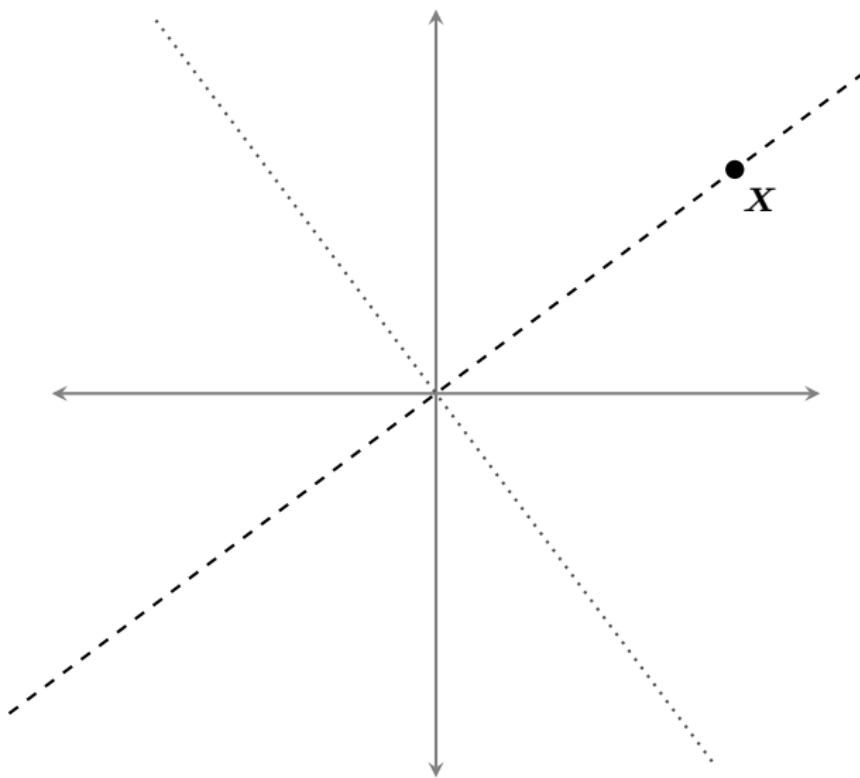


Details

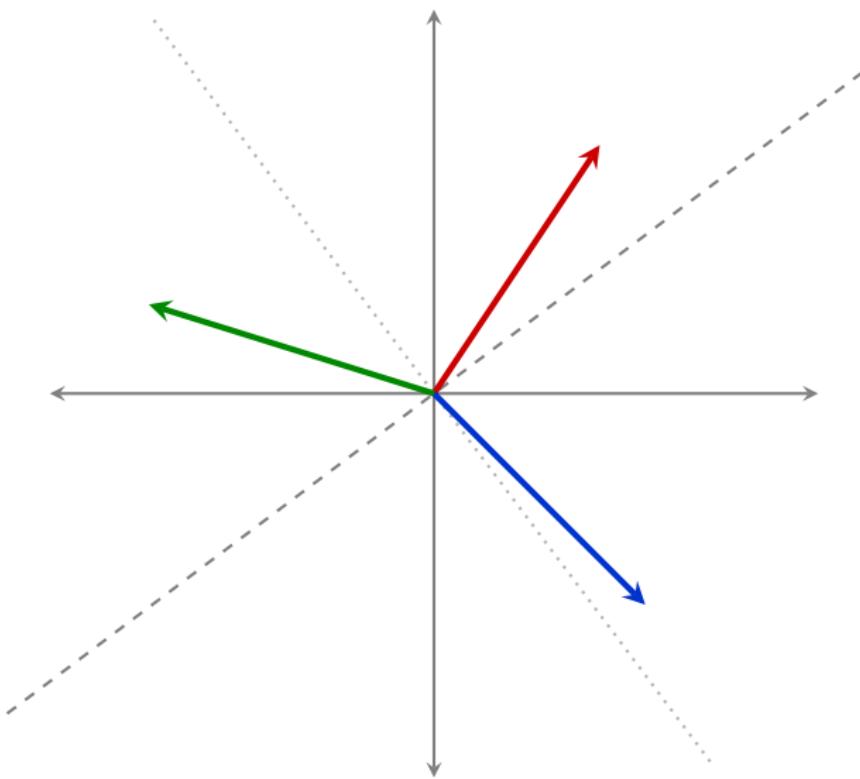
Decomposing the outcomes into two known directions



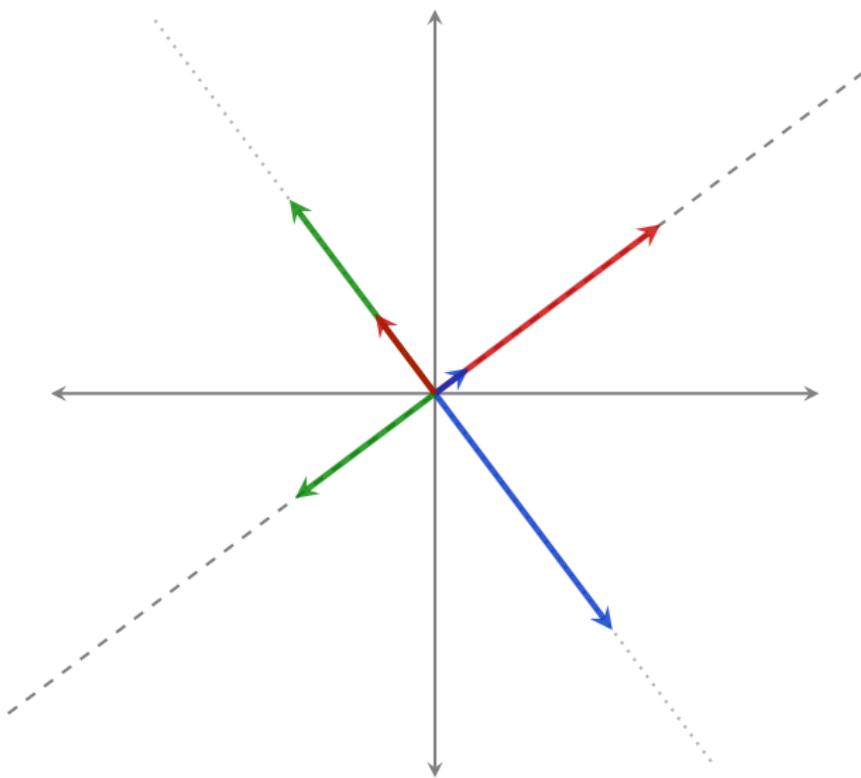
We know the axes of the decomposition



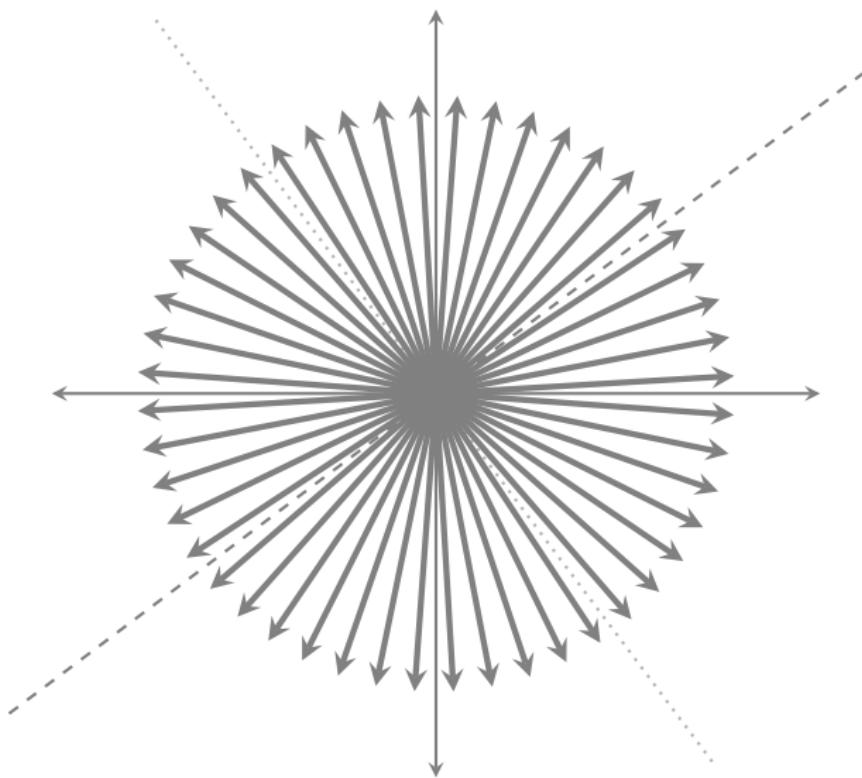
We know the axes of the decomposition



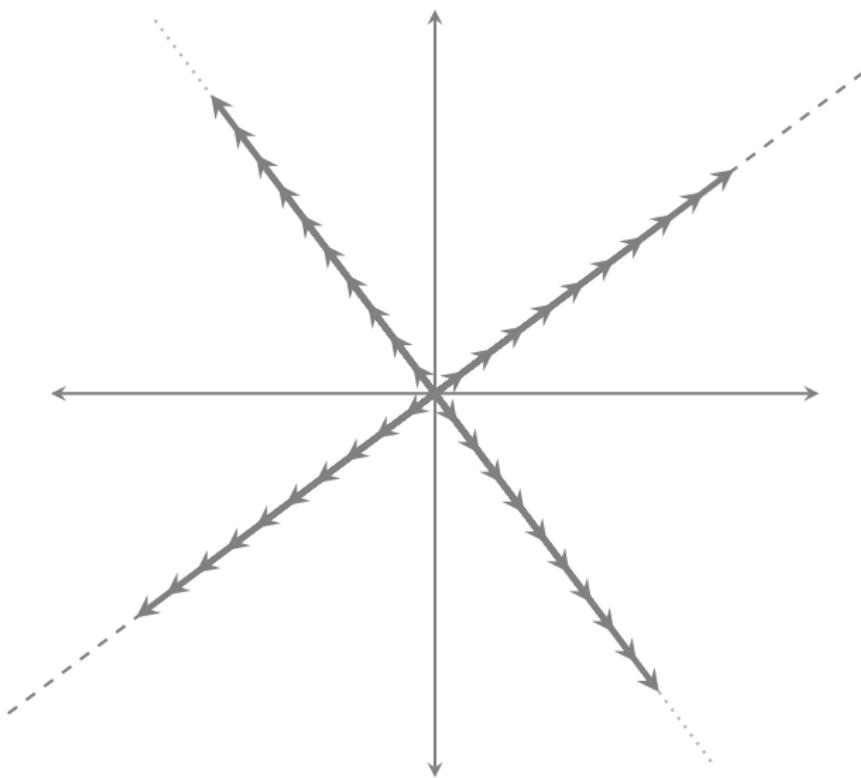
We know the axes of the decomposition



We know the axes of the decomposition



We know the axes of the decomposition

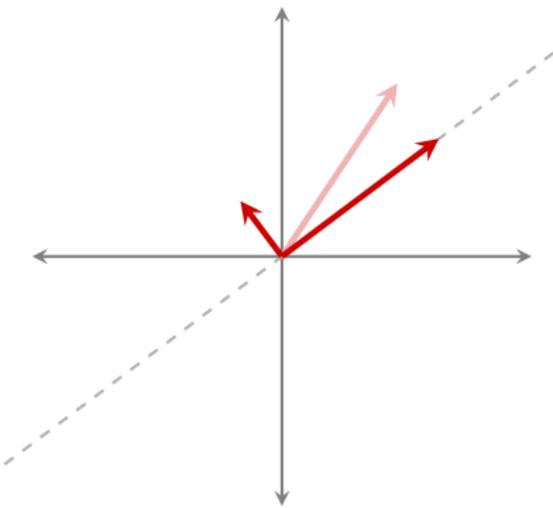


~~In what direction are the potential outcomes pointing?~~

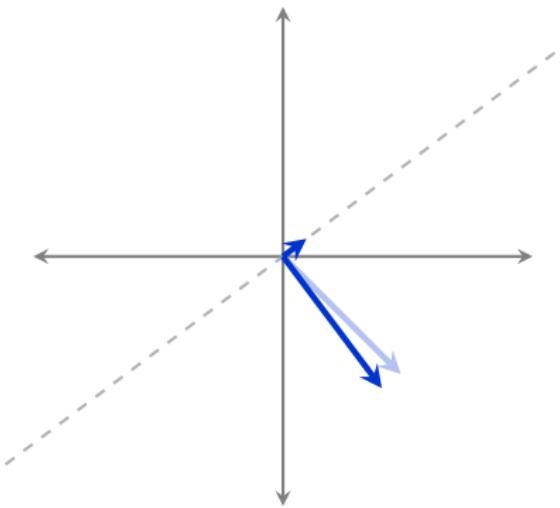
How informative are the covariates?

How informative are the covariates?

Informative



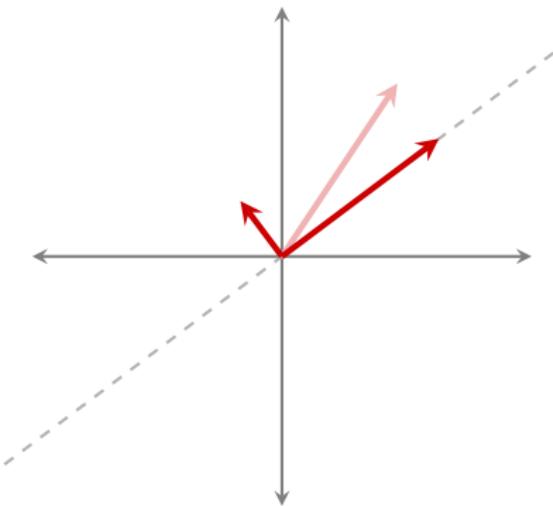
Not informative



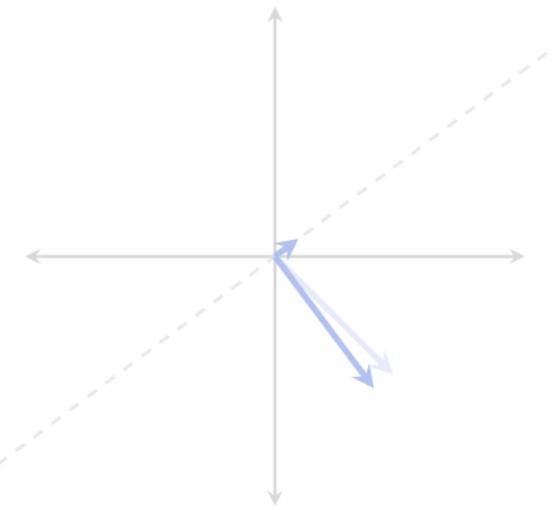
Details

How informative are the covariates?

Informative



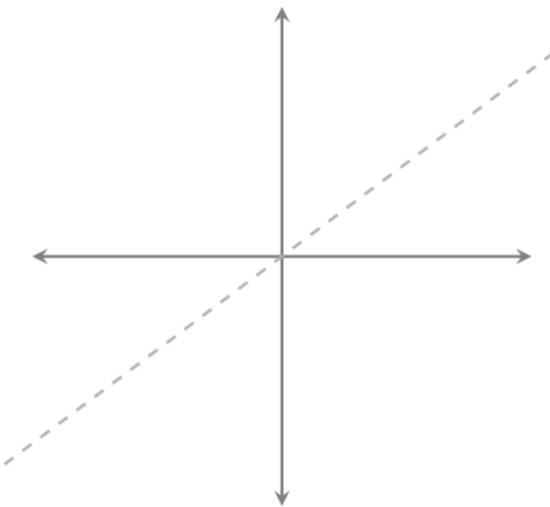
Not informative



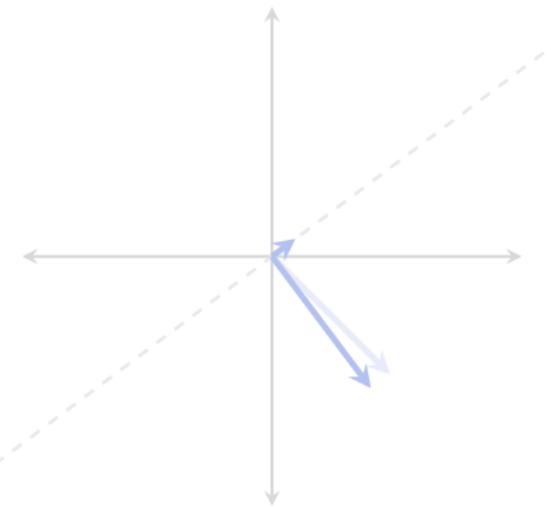
Details

How informative are the covariates?

Informative



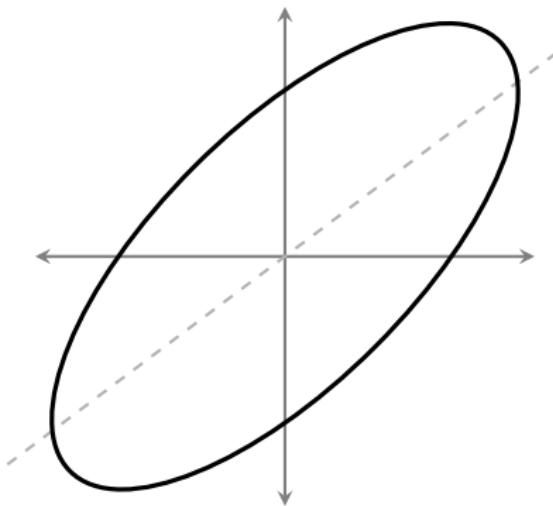
Not informative



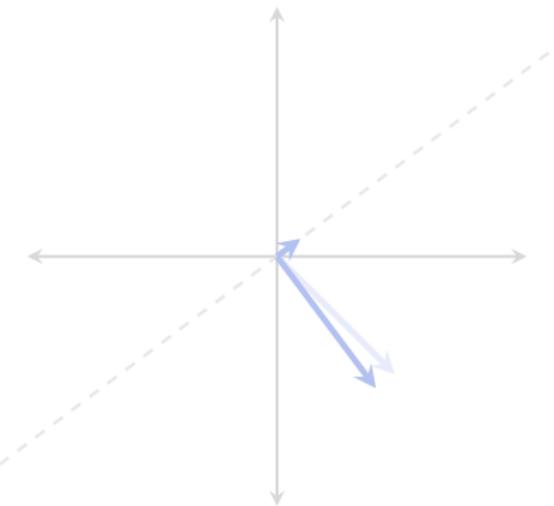
Details

How informative are the covariates?

Informative



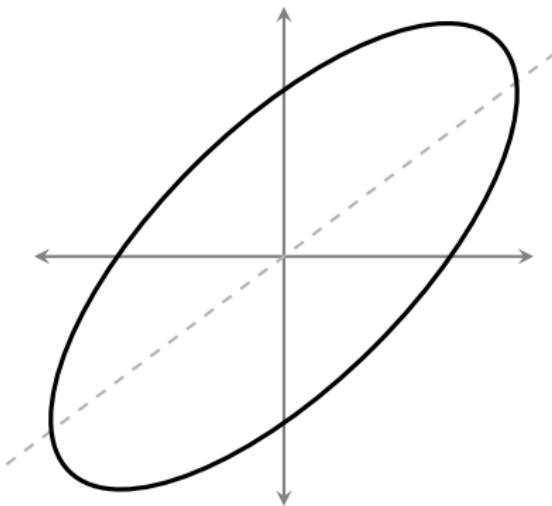
Not informative



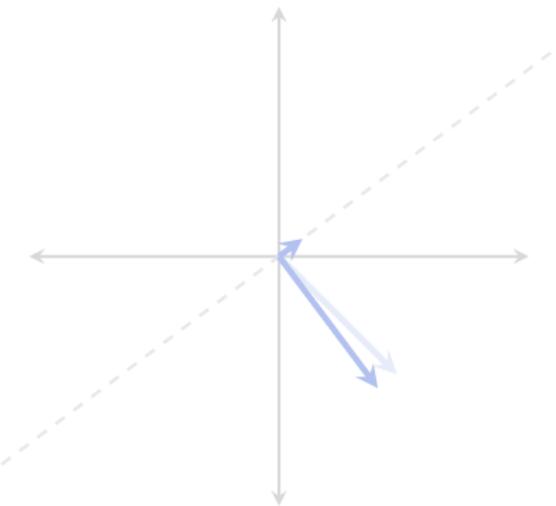
Details

How informative are the covariates?

Informative



Not informative

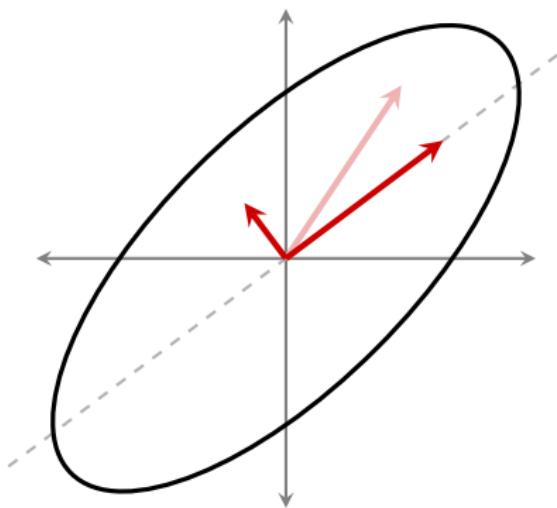


Performs well

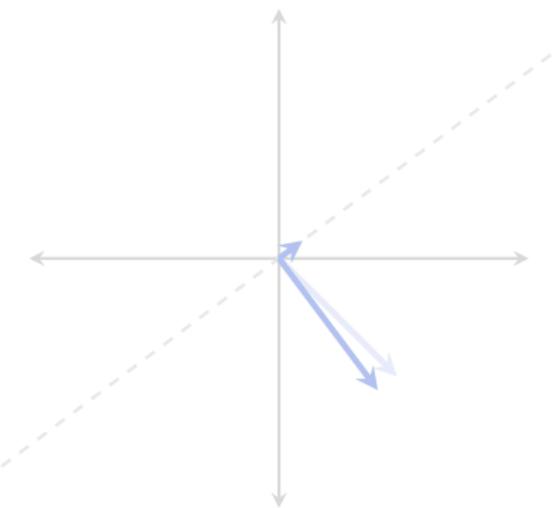
Details

How informative are the covariates?

Informative



Not informative

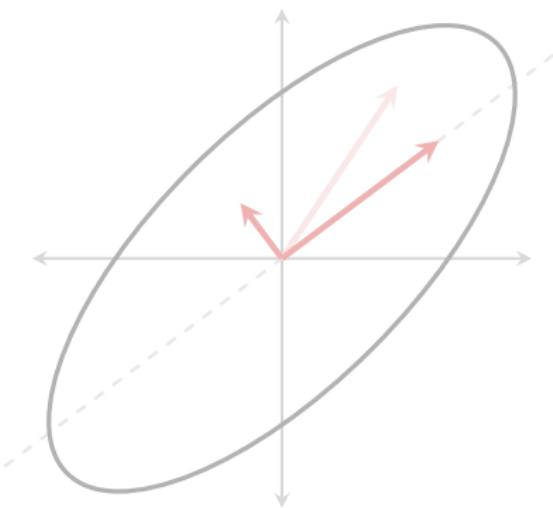


Performs well

Details

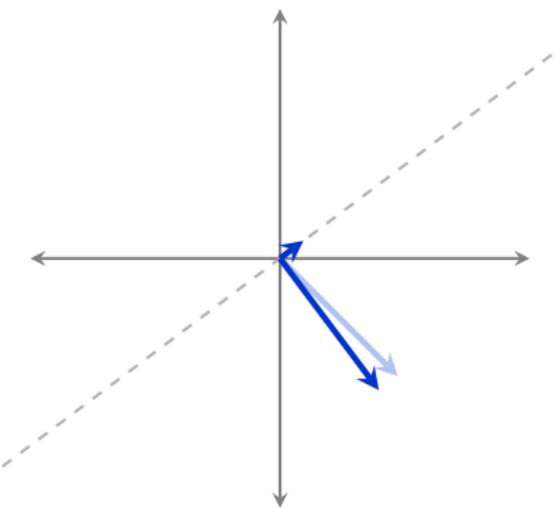
How informative are the covariates?

Informative



Performs well

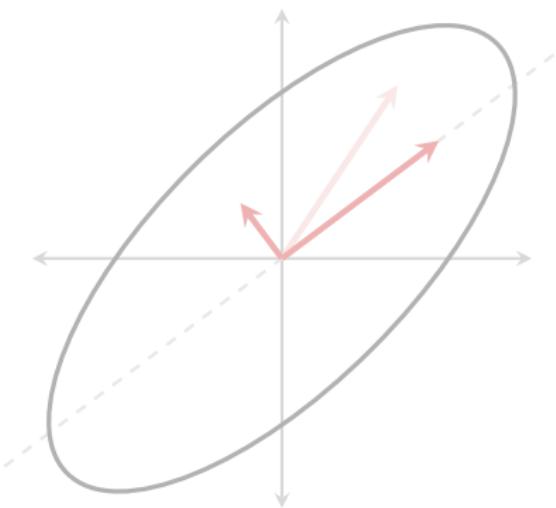
Not informative



Details

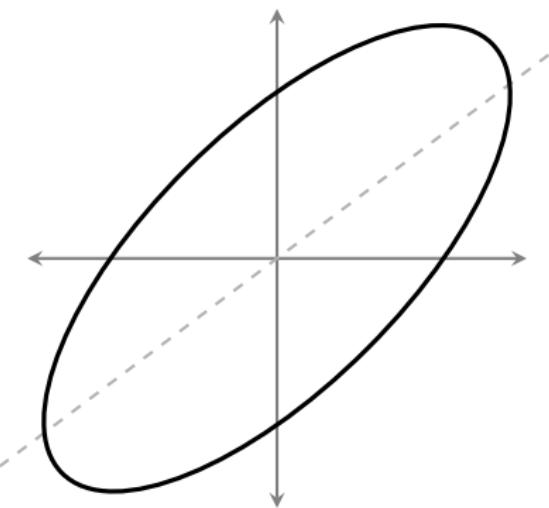
How informative are the covariates?

Informative



Performs well

Not informative

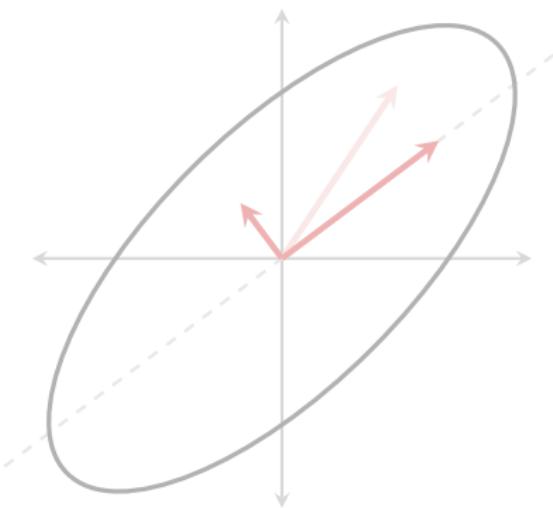


Performs poorly

Details

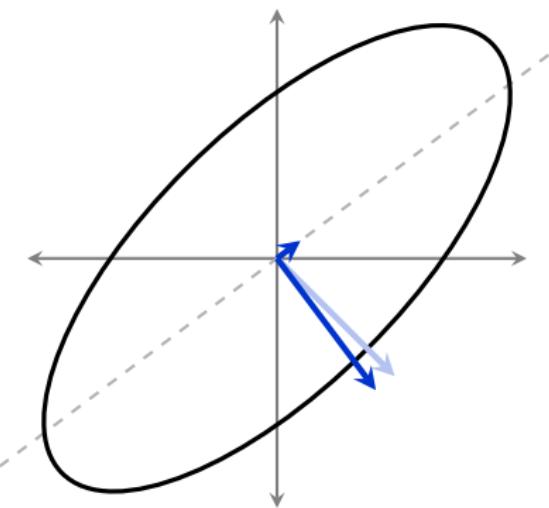
How informative are the covariates?

Informative



Performs well

Not informative

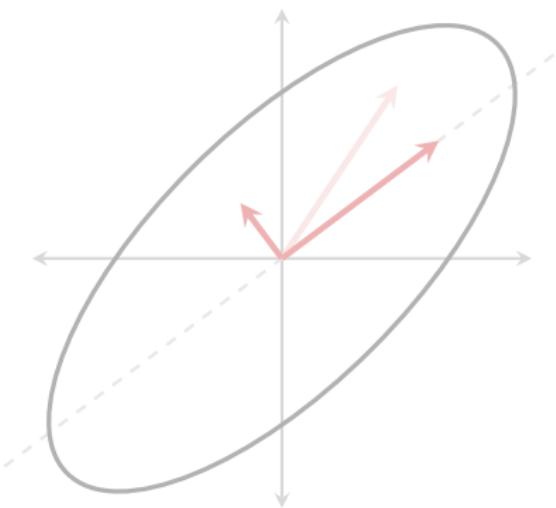


Performs poorly

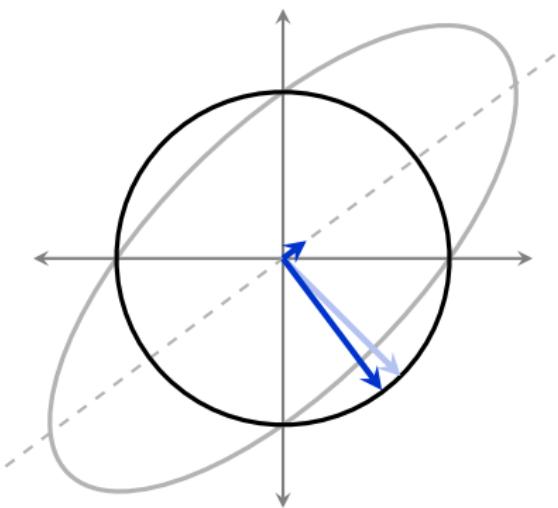
Details

How informative are the covariates?

Informative



Not informative



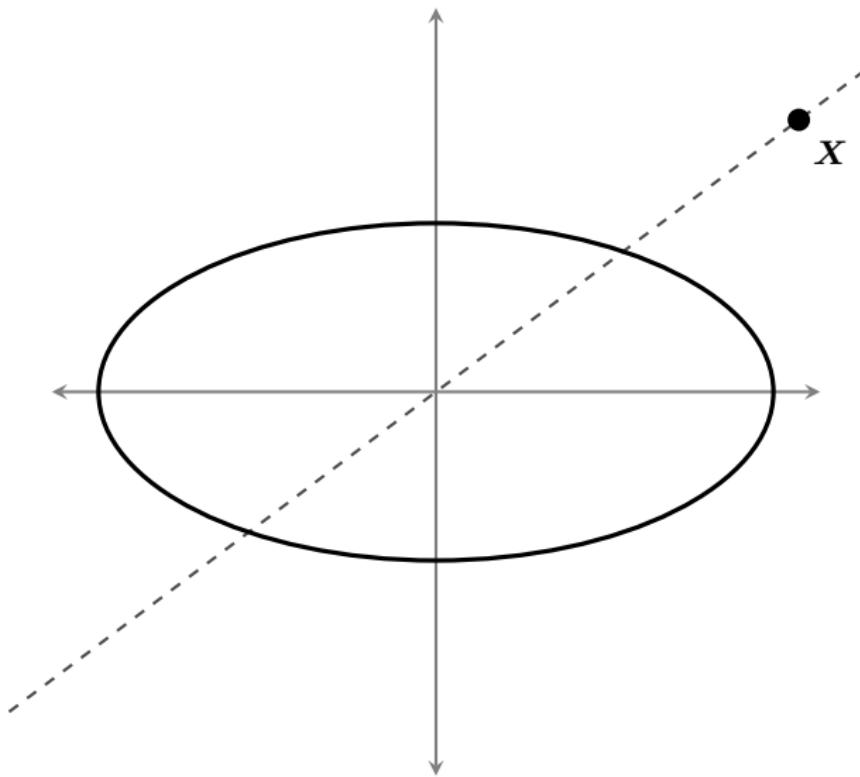
Performs well

Details

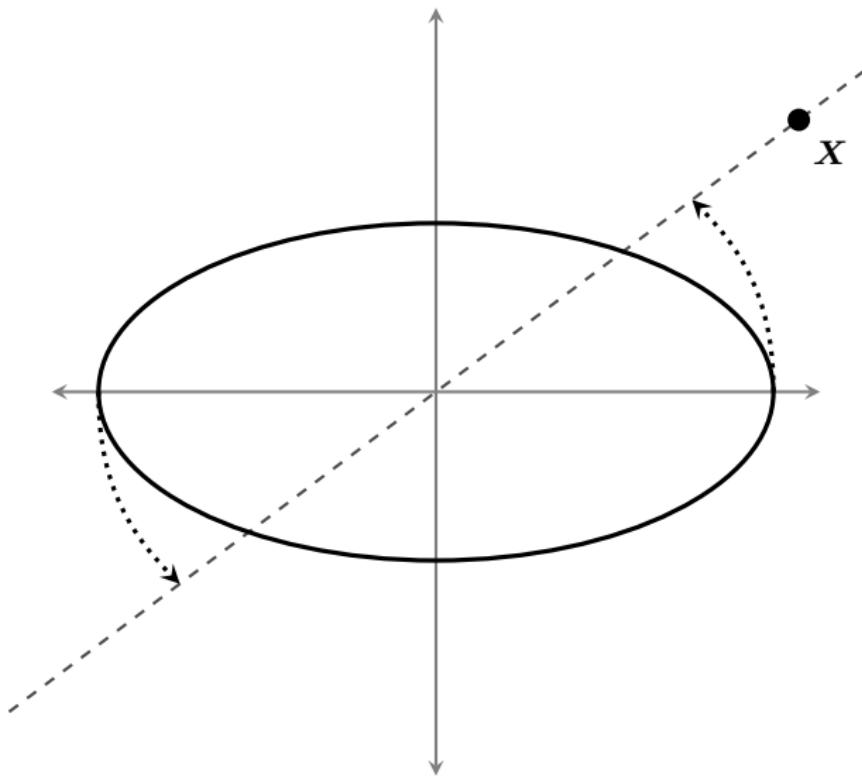
A recipe for designing experiments

- 1 Align the ellipse with the covariates.
- 2 Squeeze the ellipse to taste.
 - How informative do you believe the covariates are?
 - How much you're willing to bet on that?

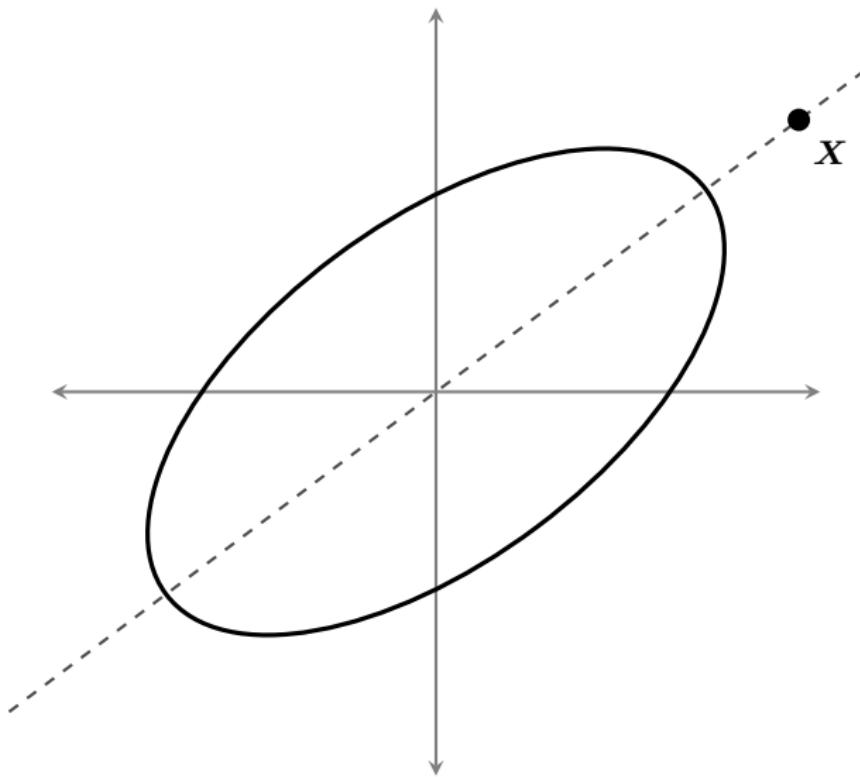
Step 1: Align the design with the covariates.



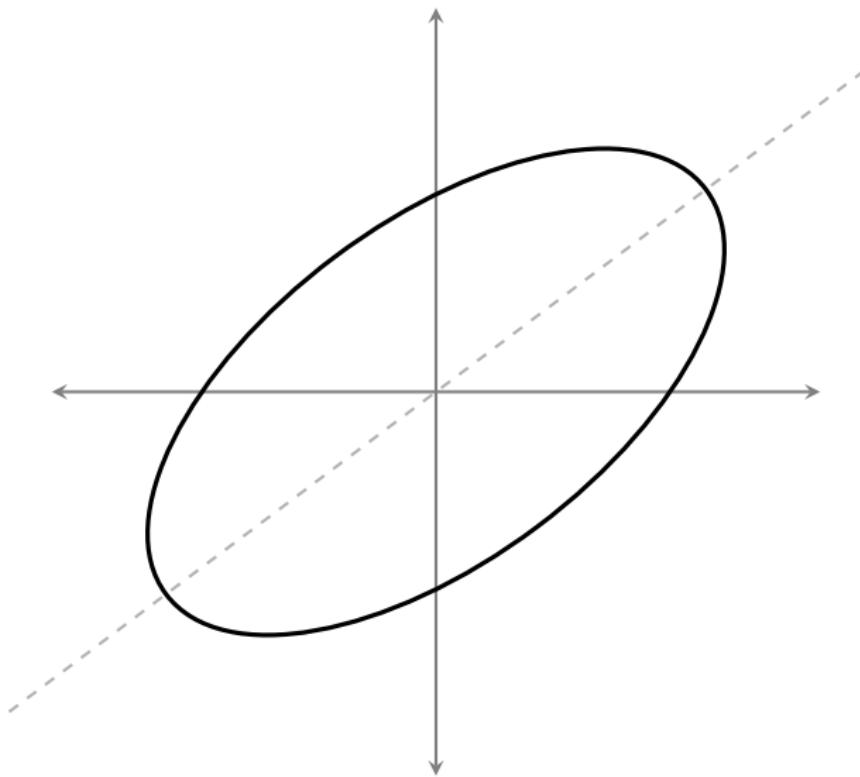
Step 1: Align the design with the covariates.



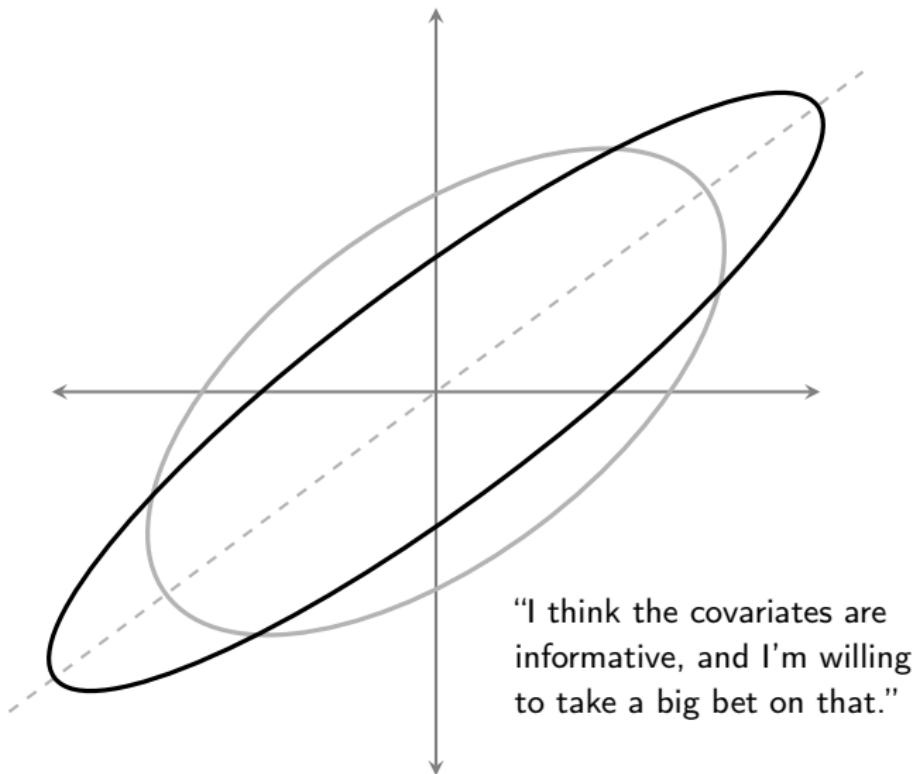
Step 1: Align the design with the covariates.



Step 2: Squeeze the ellipse to taste.

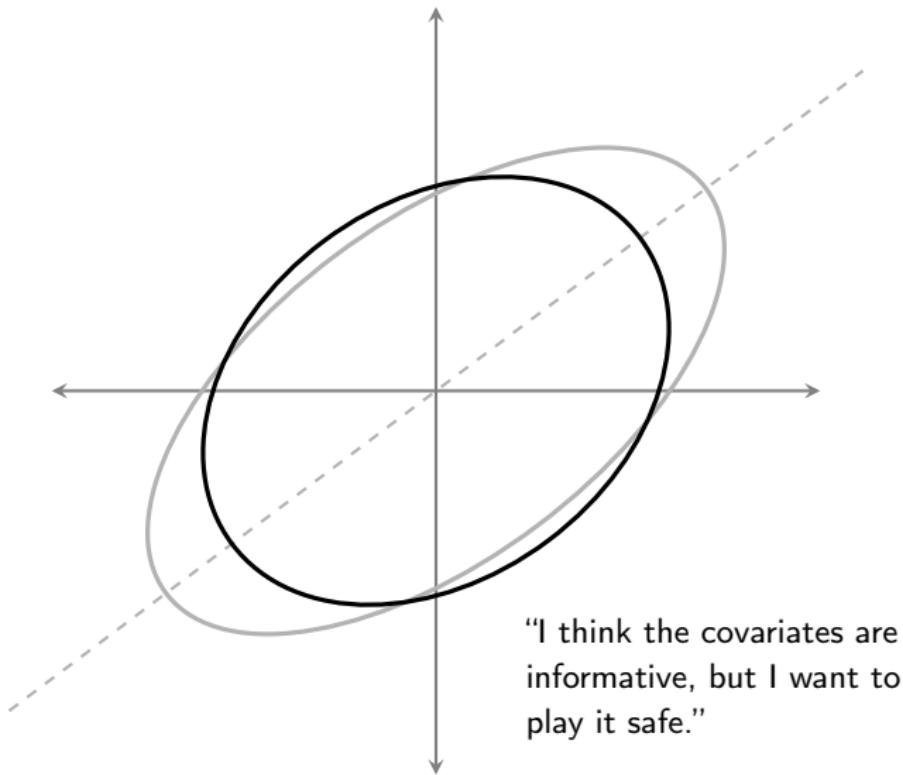


Step 2: Squeeze the ellipse to taste.



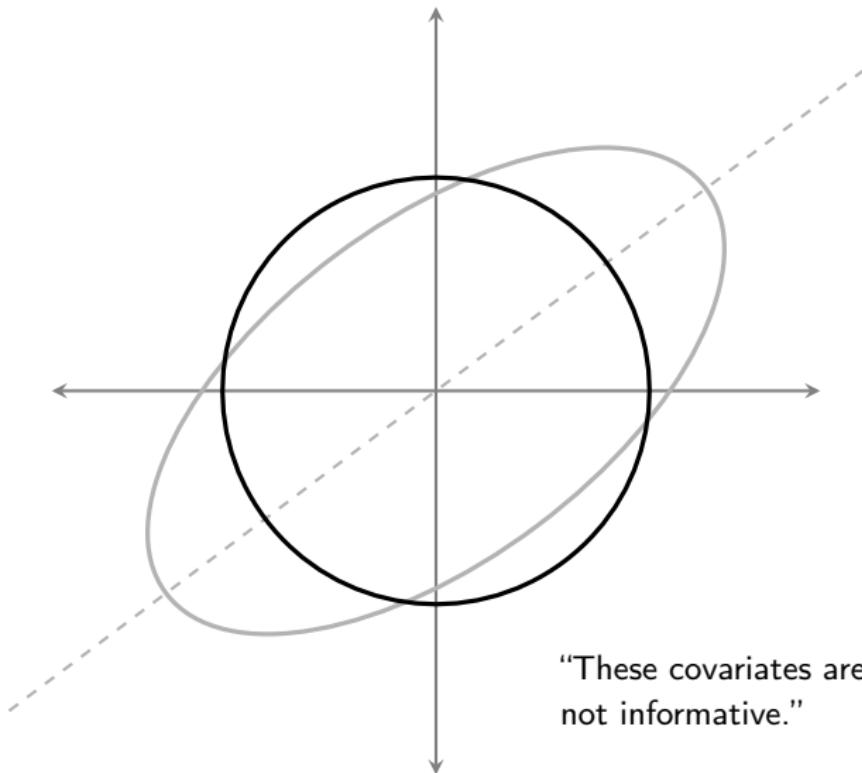
Caveat

Step 2: Squeeze the ellipse to taste.



Caveat

Step 2: Squeeze the ellipse to taste.



Caveat

Today's talk

- Why should we randomize? ✓
- How should we randomize? ✓
- How can we randomize in practice?

Today's talk

- Why should we randomize? ✓
- How should we randomize? ✓
- How can we randomize in practice?

How to do all this in practice?

Problems to tackle:

- Many ellipses (eigensystems) are not realizable.
- Even if it's realizable, unclear what design realizes it.
- Even if we know the design, unclear how to draw assignments from it.

Balancing covariates in randomized experiments using the Gram–Schmidt Walk

Christopher Harshaw

Fredrik Sävje

Daniel A. Spielman

Peng Zhang

Yale University

May 6, 2020

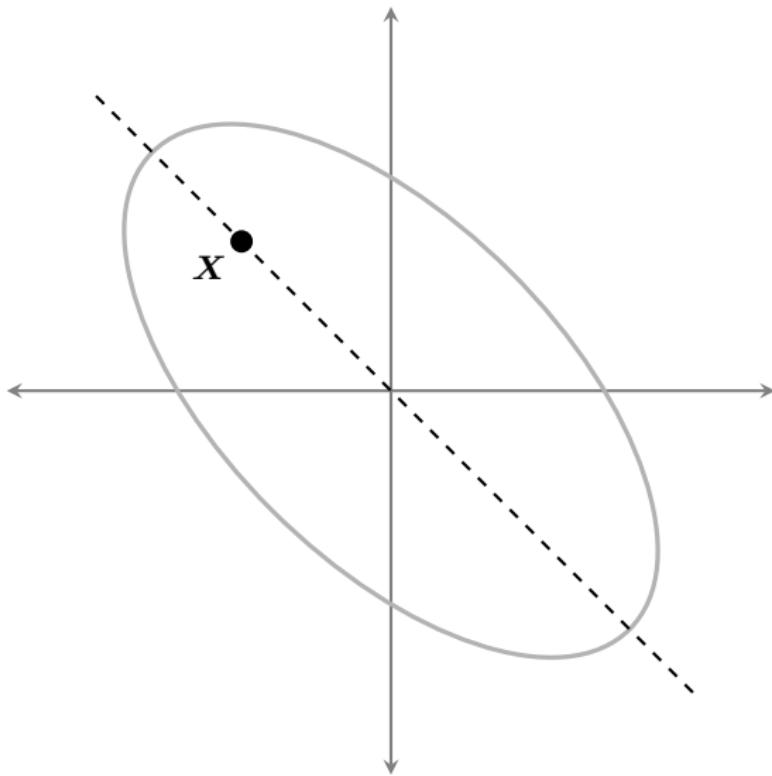
Abstract

The design of experiments involves a compromise between covariate balance and robustness. This paper introduces an experimental design that admits precise control over this trade-off. The design is specified by a parameter that bounds the worst-case mean square error of an estimator of the average treatment effect. Subject to the experimenter's desired level of robustness, the design aims to simultaneously balance all linear functions of the targeted covariates. The achieved level of balance

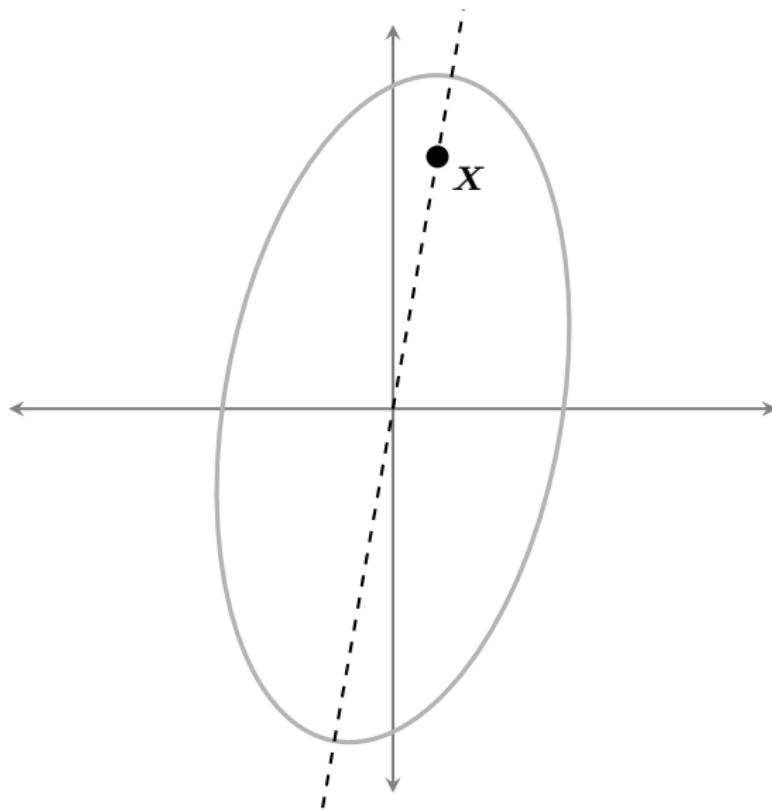
The Gram–Schmidt Walk design

- Input 1: Covariates \mathbf{X} .
 - “Please align the ellipse with this direction.”
- Input 2: Robustness parameter ϕ .
 - $\phi = 1$ prioritizes robustness: “no squeezing please.”
 - $\phi = 0$ prioritizes balance: “squeeze as much as you can.”
 - Intermediate values yield intermediate priorities.
- Output: A randomly drawn assignment vector \mathbf{Z} .

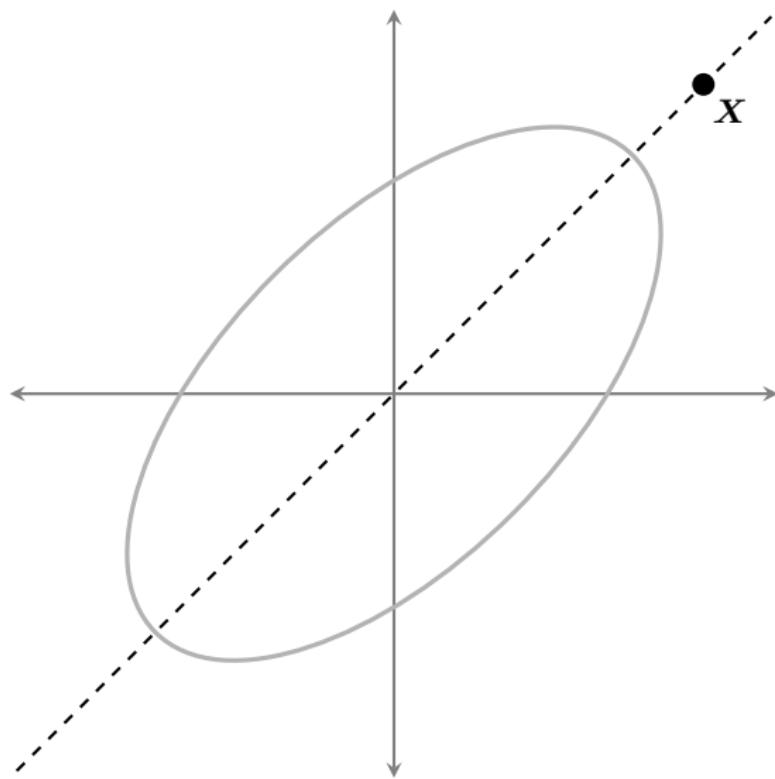
Input 1: Covariates.



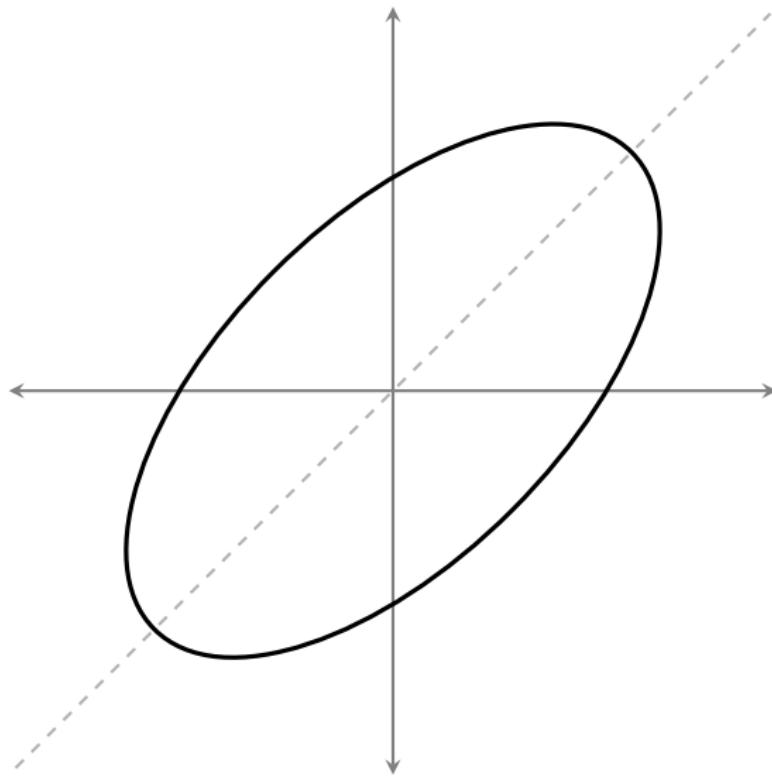
Input 1: Covariates.



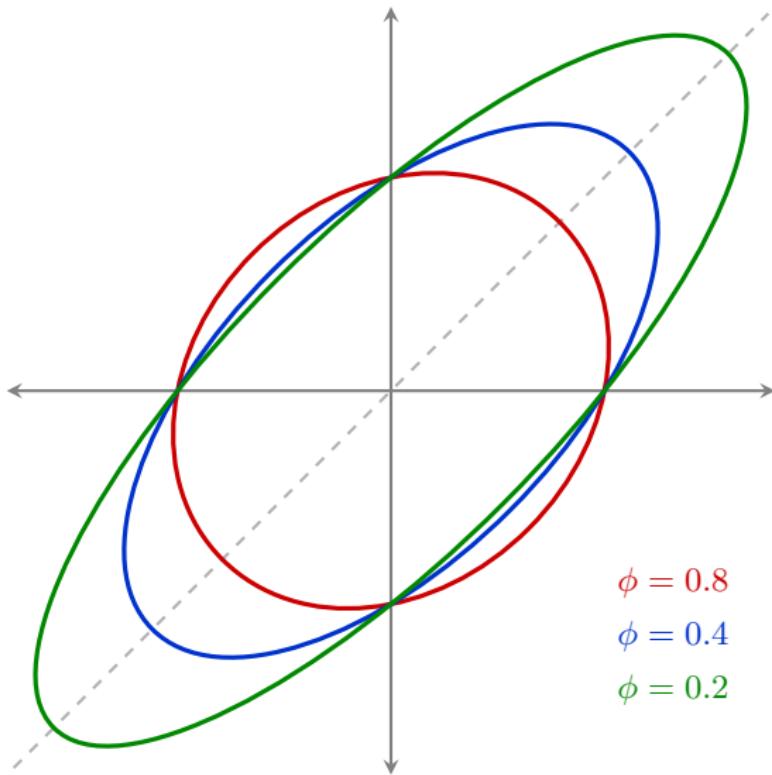
Input 1: Covariates.



Input 2: Robustness parameter.



Input 2: Robustness parameter.



RStudio

example.R x

Source on Save | Run | Source

```
1 library(gswdesign)
2
3 gswdesign_setup()
4
5 covariates <- as.matrix(read.csv("baseline.csv"))
6
7 assignments <- sample_gs_walk(covariates, phi = 0.7)
8
9 assignments
10
```

10:1 (Top Level) R Script

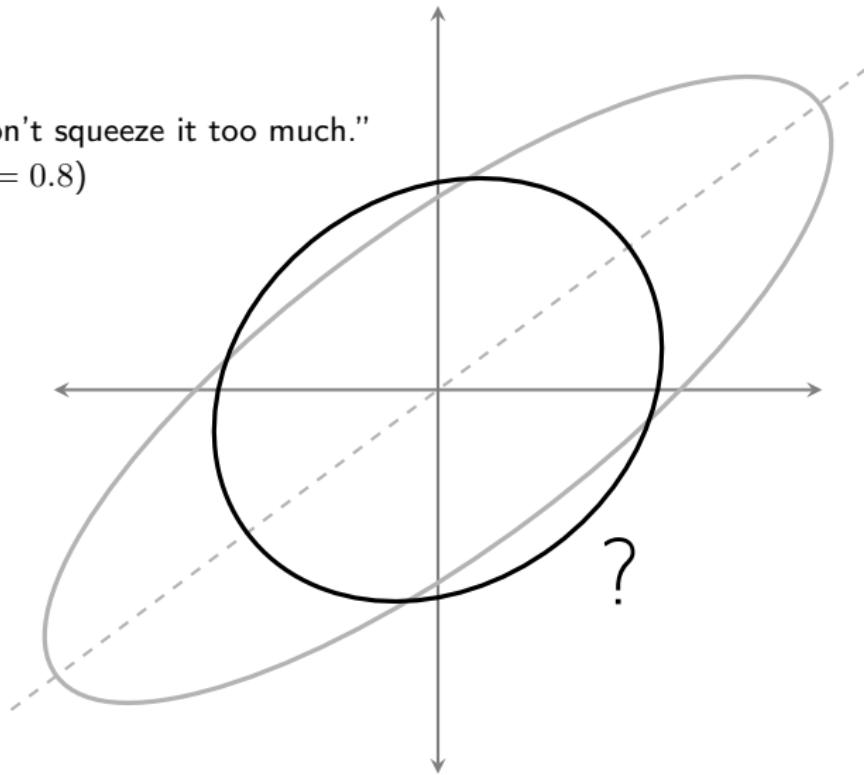
Console ~/berkeley-job-talk/

```
> library(gswdesign)
Please remember to initialize 'gswdesign' with 'gswdesign_setup()' before calling its functions.
> gswdesign_setup()
Julia version 1.5.2 will be used.
Loading setup script for JuliaCall...
Finish loading setup script for JuliaCall.
> covariates <- as.matrix(read.csv("baseline.csv"))
> assignments <- sample_gs_walk(covariates, phi = 0.7)
> assignments
[1] TRUE TRUE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE TRUE TRUE FALSE FALSE FALSE
[16] TRUE FALSE TRUE FALSE TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE TRUE
[31] FALSE TRUE FALSE TRUE FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE TRUE TRUE TRUE
[46] FALSE FALSE FALSE TRUE TRUE FALSE FALSE TRUE TRUE FALSE TRUE TRUE FALSE FALSE FALSE
```

Does it work?

Does ϕ actually control how squeezed the ellipse is?

"Don't squeeze it too much."
 $(\phi = 0.8)$

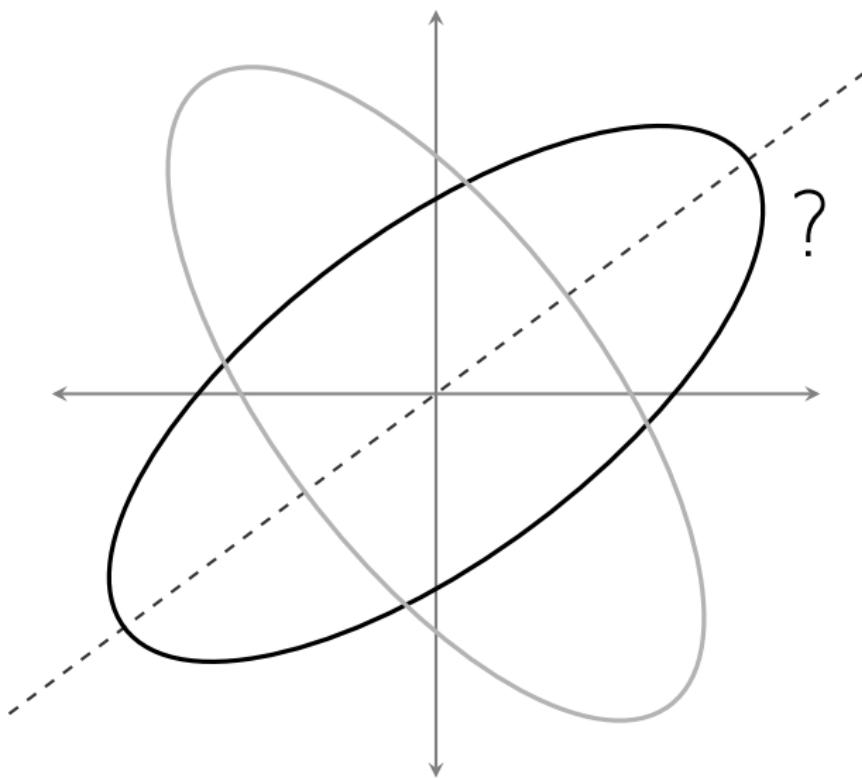


Theorem

In the worst-case, when covariates are completely uninformative, the standard error under the Gram–Schmidt Walk design is at most $100/\sqrt{\phi}$ percent of the error under independent assignment.

- $\phi = 1$: no difference.
- $\phi = 0.9$: at most 5.4% higher.
- $\phi = 0.8$: at most 11.8% higher.
- $\phi = 0.5$: at most 41.4% higher (same as matched pair design).

Does the GSW design actually align the ellipse with the covariates?



Theorem

The covariate imbalance under the Gram–Schmidt Walk design is at least $100 - 100 / \sqrt{\phi + (1 - \phi) / \gamma}$ percent lower than under independent assignment.

- γ gives the theoretical limit for how aligned the design can be with the covariates.
 - Roughly, $\gamma \approx \{\#\text{covariates}\} \times \log(n) / n$.
- With 100 units and 4 covariates, the decreases in imbalance are:
 - $\phi = 0.9$: at least 16.7%.
 - $\phi = 0.8$: at least 27.2%.
 - $\phi = 0.5$: at least 44.2%.

*Simplification using decorrelated covariates.

Does it work?

Yes

(See paper for many additional results.)

Today's talk

- Why should we randomize? ✓
- How should we randomize? ✓
- How can we randomize in practice? ✓

Main takeaways

- Don't gamble with your experiments. Please randomize!
- If you have a subset of informative covariates, it's generally a good idea to "pay" to have them balanced.
- But don't pay too much. Rule of thumb: $\phi \geq 0.5$.
 - Covariates are generally not extremely informative.
 - The price for balance is non-linear.
- These insights apply to all designs.
 - The Gram–Schmidt Walk design makes it easy to control the trade-off.

Thanks!

<https://fredriksavje.com/slides.pdf>

How does it work?

Simulation results

Regression by design

Confidence intervals

Consider the treatment effect estimator

$$\hat{\tau} = \frac{1}{n/2} \sum_{i=1}^n Z_i Y_i(1) - \frac{1}{n/2} \sum_{i=1}^n (1 - Z_i) Y_i(0).$$

Let's focus on the contribution of unit i :

$$\frac{Z_i Y_i(1)}{n/2} - \frac{(1 - Z_i) Y_i(0)}{n/2}.$$

We can rewrite this as

$$\frac{\tau_i}{n} + \frac{(2Z_i - 1)\mu_i}{n/2}.$$

Show derivation

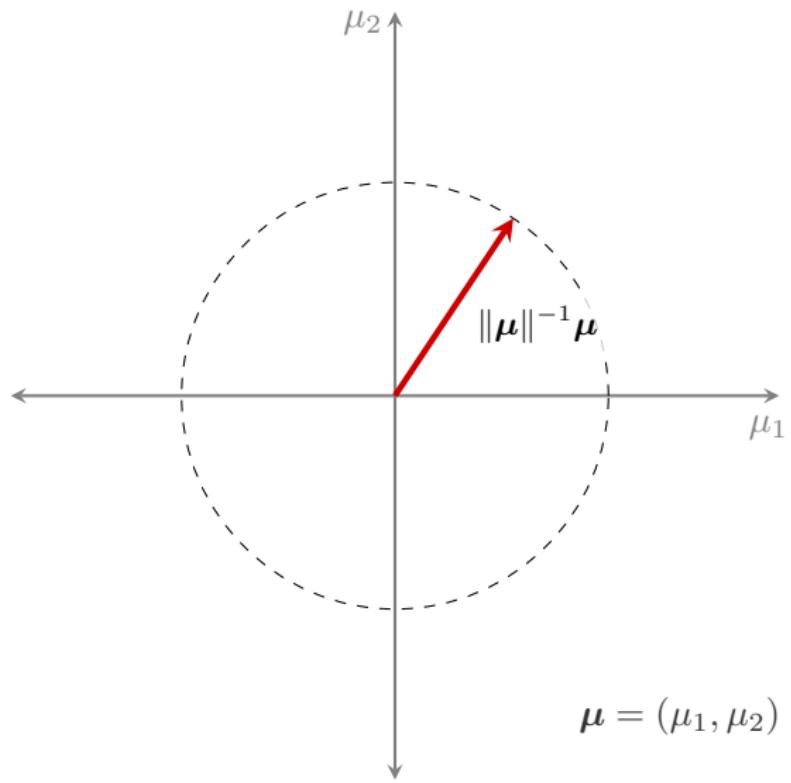
If $\mu_i = 0$, then unit i 's contribution to the estimator is fixed at τ_i/n .

- This is unit i 's treatment effect, so it doesn't contribute to bias.
- It is not random, so the unit does not contribute to variance.

Back

$$\begin{aligned}
& \frac{Z_i Y_i(1)}{n/2} & - & \frac{(1 - Z_i) Y_i(0)}{n/2} \\
& \frac{Z_i [Y_i(1) + Y_i(0)]}{n/2} & - & \frac{Y_i(0)}{n/2} \\
& \frac{2Z_i \mu_i}{n/2} & - & \frac{2Y_i(0)}{n} \\
& \frac{2Z_i \mu_i}{n/2} - \frac{\mu_i}{n/2} & + & \frac{2\mu_i}{n} - \frac{2Y_i(0)}{n} \\
& \frac{2Z_i \mu_i - \mu_i}{n/2} & + & \frac{Y_i(1) + Y_i(0)}{n} - \frac{2Y_i(0)}{n} \\
& \frac{(2Z_i - 1)\mu_i}{n/2} & + & \frac{Y_i(1) - Y_i(0)}{n} \\
& \frac{(2Z_i - 1)\mu_i}{n/2} & + & \frac{\tau_i}{n}
\end{aligned}$$

Normalized mean potential outcome vector



Let $\boldsymbol{\eta}$ denote an eigenvector and λ denote an eigenvalue. That is, let $\boldsymbol{\eta}$ and λ be such that

$$\text{Corr}(\mathbf{Z})\boldsymbol{\eta} = \lambda\boldsymbol{\eta}.$$

In other words, $\boldsymbol{\eta}$ is direction invariant when transformed by $\text{Corr}(\mathbf{Z})$.

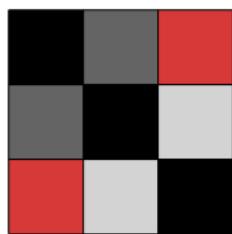
When $n = 2$, two solutions exist to this equation. Sort them by the size of the eigenvalue:

$$\boldsymbol{\eta}_{\min}, \lambda_{\min} \quad \boldsymbol{\eta}_{\max}, \lambda_{\max}$$

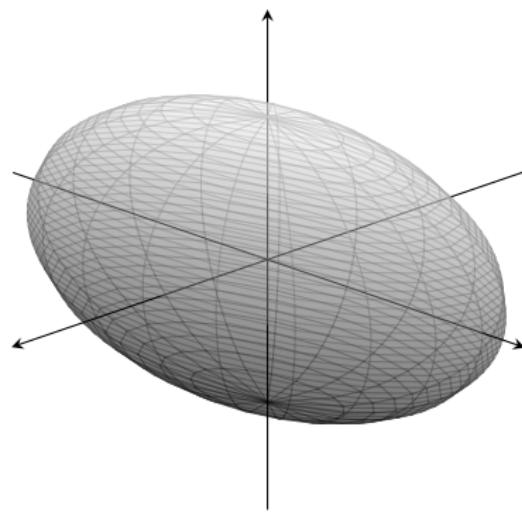
Because covariance matrices are symmetric, the eigenvectors are orthogonal.
Because covariance matrices are positive semidefinite, all eigenvalues are non-negative real values.

Thus the eigensystem can be interpreted as the directions and magnitudes of the axes of an ellipse ($n = 2$), ellipsoid ($n = 3$) or hyperellipsoid ($n \geq 4$).

Experimental designs as ellipsoids when $n = 3$



=



Spectral decomposition of the estimation error

$$E[(\hat{\tau} - \tau)^2] \propto \sum_{i=1}^n w_i^2 \lambda_i \quad \text{where} \quad w_i = \langle \mu, \eta_i \rangle.$$

- Proportionality constant does not depend on the design.
- Symbol legend:
 - μ : Vector of mean potential outcomes (normalized).
 - η_i : The i th eigenvector of the design's correlation matrix.
 - λ_i : The i th eigenvalue of the design's correlation matrix.
 - w_i : Alignment of the potential outcomes with the i th eigenvector.

We have $E[(\hat{\tau} - \tau)^2] \propto \boldsymbol{\mu}^\top \text{Corr}(\mathbf{Z}) \boldsymbol{\mu}$.

Because $\text{Corr}(\mathbf{Z})$ is symmetric, its eigenvectors $\boldsymbol{\eta}_i$ form an orthonormal basis of the full \mathbb{R}^n vector space. We can decompose any $\boldsymbol{\mu} \in \mathbb{R}^n$ with $\|\boldsymbol{\mu}\| = 1$ as

$$\boldsymbol{\mu} = \sum_{i=1}^n w_i \boldsymbol{\eta}_i \quad \text{where} \quad w_i = \langle \boldsymbol{\mu}, \boldsymbol{\eta}_i \rangle.$$

This means that we can write

$$\boldsymbol{\mu}^\top \text{Corr}(\mathbf{Z}) \boldsymbol{\mu} = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \boldsymbol{\eta}_i^\top \text{Corr}(\mathbf{Z}) \boldsymbol{\eta}_j.$$

By the definition of eigenvectors, we have $\text{Corr}(\mathbf{Z}) \boldsymbol{\eta}_j = \lambda_j \boldsymbol{\eta}_j$. Because the eigenvectors are orthonormal, $\boldsymbol{\eta}_i^\top \boldsymbol{\eta}_i = 1$ and $\boldsymbol{\eta}_i^\top \boldsymbol{\eta}_j = 0$. Thus

$$\boldsymbol{\mu}^\top \text{Corr}(\mathbf{Z}) \boldsymbol{\mu} = \sum_{i=1}^n w_i^2 \lambda_i.$$

Recall the estimation error:

$$\mathrm{E}[(\hat{\tau} - \tau)^2] \propto \sum_{i=1}^n w_i^2 \lambda_i \quad \text{where} \quad w_i = \langle \boldsymbol{\mu}, \boldsymbol{\eta}_i \rangle.$$

Under independent assignment, $\mathrm{Corr}(\mathbf{Z}) = \mathbf{I}$, so $\lambda_1 = \dots = \lambda_n = 1$.

By construction, $w_1^2 + \dots + w_n^2 = 1$. Hence,

$$\sum_{i=1}^n w_i^2 \lambda_i = 1 \quad \text{for all } \boldsymbol{\mu} \in \mathbb{R}^n.$$

Note that $\text{Corr}(Z_i, Z_i)$ is fixed at 1. Hence, the trace of $\text{Corr}(\mathbf{Z})$ is fixed at n for all designs.

Recall that the sum of the eigenvalues of a matrix is equal to its trace:

$$\text{tr}(\text{Corr}(\mathbf{Z})) = \sum_{i=1}^n \lambda_i.$$

Therefore, the only way to have $\lambda_{\min} < 1$ is to accept $\lambda_{\max} > 1$.

If a design potentially could perform better than independent assignment, it must have $\lambda_{\min} < 1$.

This implies that $\lambda_{\max} > 1$ for the design.

Consider the potential outcomes $\mu = \eta_{\max}$.

Independent assignment has $E[(\hat{\tau} - \tau)^2] \propto 1$, as always.

The alternative design has $E[(\hat{\tau} - \tau)^2] \propto \lambda_{\max} > 1$, with the same proportionality constant as above.

When a design is (semi-)deterministic (i.e., one coin flip for all units), the correlation matrix consists of only 1 (perfectly positively correlated) or -1 (perfectly negatively correlated).

The eigenvalues for such a matrix are always

$$\lambda_{\max} = n \quad \text{and} \quad \lambda_1 = \dots = \lambda_{n-1} = 0.$$

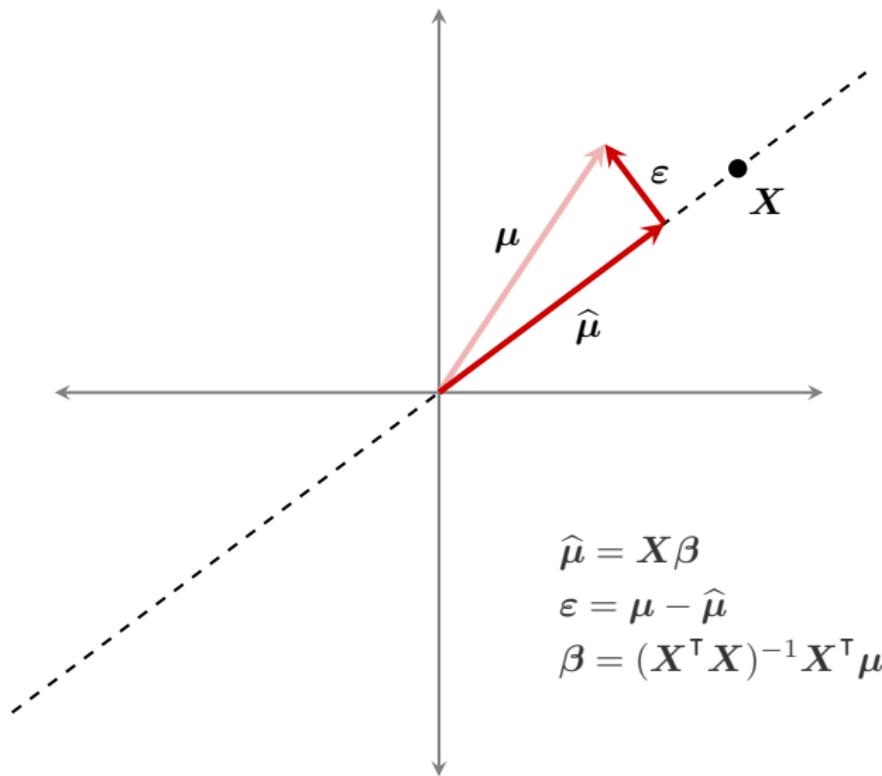
Hence, the performance of a deterministic design is

$$E[(\hat{\tau} - \tau)^2] \propto w_{\max}^2 n = \langle \boldsymbol{\mu}, \boldsymbol{\eta}_{\max} \rangle^2 n.$$

Theorem

Provided that the second moment of the mean potential outcome vector is asymptotically bounded, the average treatment effect estimator is consistent if and only if the maximum eigenvalue of the design's covariance matrix is asymptotically dominated by the sample size.

Orthogonal projection onto covariate subspace



We have $E[(\hat{\tau} - \tau)^2] \propto \mu^\top \text{Corr}(Z)\mu$.

Orthogonal projection of μ onto covariate subspace:

$$\hat{\mu} = X\beta \quad \varepsilon = \mu - \hat{\mu} \quad \beta = (X^\top X)^{-1} X^\top \mu$$

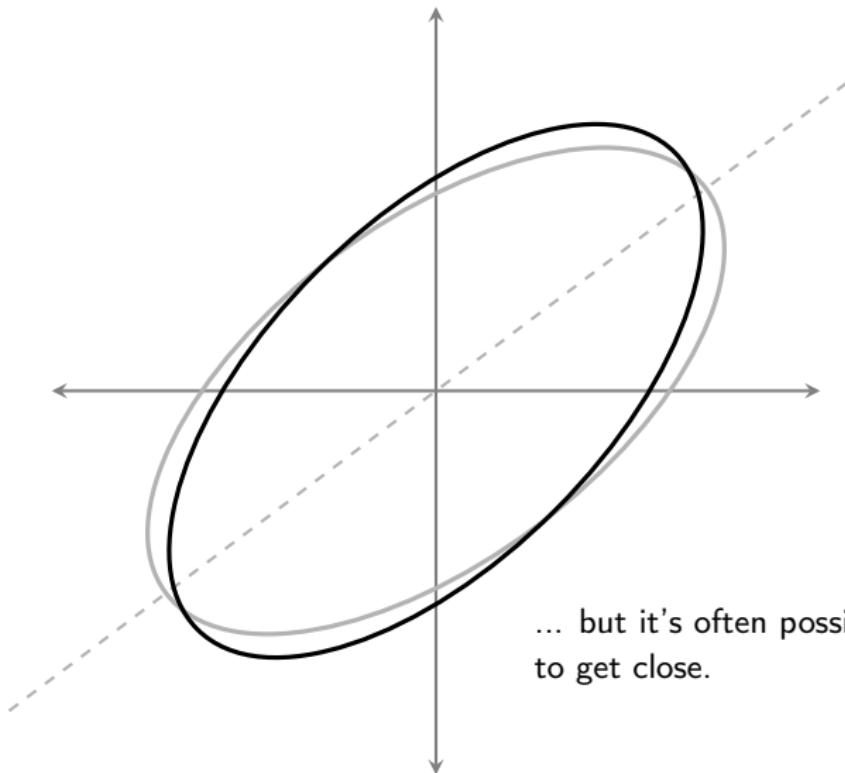
Because $\mu = \hat{\mu} + \varepsilon$,

$$\mu^\top \text{Corr}(Z)\mu = \hat{\mu}^\top \text{Corr}(Z)\hat{\mu} + \varepsilon^\top \text{Corr}(Z)\varepsilon + 2\hat{\mu}^\top \text{Corr}(Z)\varepsilon$$

Just as above, we can use the spectral interpretation to understand

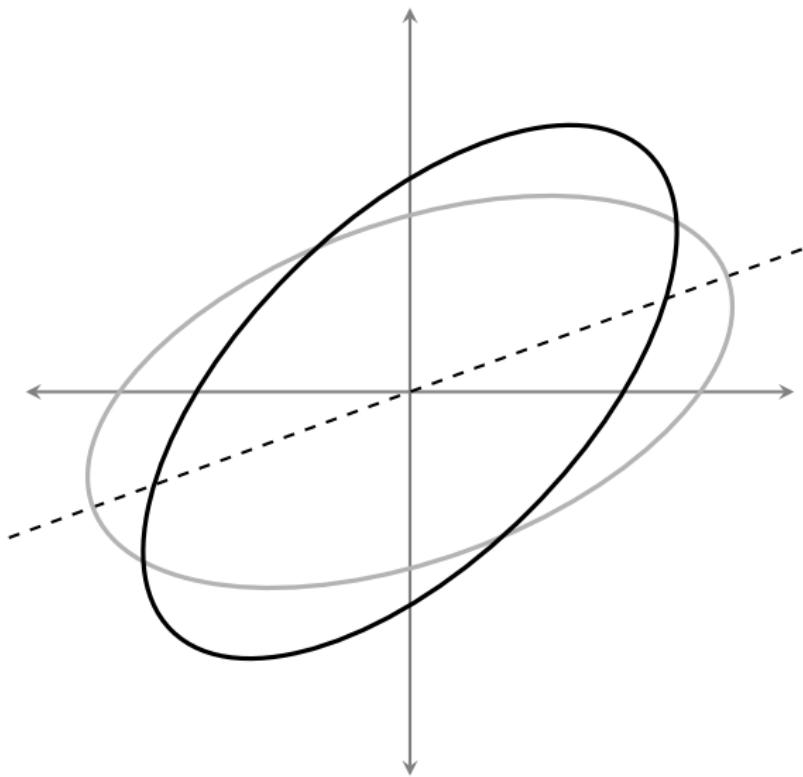
$$\hat{\mu}^\top \text{Corr}(Z)\hat{\mu} \quad \text{and} \quad \varepsilon^\top \text{Corr}(Z)\varepsilon.$$

It's not always possible to perfectly align the design...



... but it's often possible
to get close.

Best realizable alignment.



Theorem

The worst-case mean squared error under the Gram–Schmidt Walk design is upper bounded by the ratio between the minimax optimum (independent assignment) and the design parameter. That is, for all potential outcome vectors μ , all covariate matrices X , and all parameter values $\phi \in (0, 1]$,

$$E_{gsw}[(\hat{\tau} - \tau)^2] \leq \frac{1}{\phi} E_{ind}[(\hat{\tau} - \tau)^2].$$

$E[(\boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{Z})^2]$ is the balance (in mean square sense) of the linear function $\boldsymbol{\theta}$ of the covariates \mathbf{X} .

Theorem

For all covariate matrices \mathbf{X} , all functions $\boldsymbol{\theta}$, and all parameters $\phi \in (0, 1]$,

$$\frac{E_{gsw}[(\boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{Z})^2]}{E_{ind}[(\boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{Z})^2]} \leq \left(\phi \lambda_G^{-1} + (1 - \phi) \xi^{-2} \right)^{-1} \frac{\|\boldsymbol{\theta}\|^2}{\|\mathbf{X}\boldsymbol{\theta}\|^2},$$

where λ_G is the maximum eigenvalue of the Gram matrix $\mathbf{X}^\top \mathbf{X}$, and $\xi = \max_{i \in \{1, \dots, n\}} \|\mathbf{x}_i\|$ is the maximum row norm of the covariate matrix.

For well-behaved covariates, $\lambda_G \approx n$ and $\xi^2 \approx \{\#\text{covariates}\} \times \log(n)$.

For relevant functions, $\|\boldsymbol{\theta}\|^2 / \|\mathbf{X}\boldsymbol{\theta}\|^2 \approx 1/n$.

Theorem

For all covariate matrices \mathbf{X} , all functions θ , and all parameters $\phi \in (0, 1]$,

$$\frac{\mathbb{E}_{gsw}[(\theta^\top \mathbf{X}^\top \mathbf{Z})^2]}{\mathbb{E}_{ind}[(\theta^\top \mathbf{X}^\top \mathbf{Z})^2]} \leq \left(\phi \lambda_G^{-1} + (1 - \phi) \xi^{-2} \right)^{-1} \frac{\|\theta\|^2}{\|\mathbf{X}\theta\|^2},$$

where λ_G is the maximum eigenvalue of the Gram matrix $\mathbf{X}^\top \mathbf{X}$, and
 $\xi = \max_{i \in \{1, \dots, n\}} \|\mathbf{x}_i\|$ is the maximum row norm of the covariate matrix.

If the covariates are decorrelated, then $\mathbf{X}^\top \mathbf{X} = n\mathbf{I}$, and

$$\lambda_G = n \quad \text{and} \quad \|\theta\|^2 / \|\mathbf{X}\theta\|^2 = 1/n.$$

To get the theorem in the main presentation, let $\gamma = \xi^2/n$ and rearrange the right-hand side.



Understanding and misunderstanding randomized controlled trials

Angus Deaton^{a,b,c,*}, Nancy Cartwright^{d,e}

^a Princeton University, USA

^b National Bureau of Economic Research, USA

^c University of Southern California, USA

^d Durham University, England

^e UC San Diego, USA

ARTICLE INFO

Keywords:

RCTs

Randomness

Contrary to frequent claims in the applied literature, randomization does *not* equalize everything other than the treatment in the treatment and control groups, it does not automatically deliver a precise estimate of the average treatment effect (ATE), and it does not relieve us of the need to think about (observed or unobserved) covariates.

Economic development

ABSTRACT

Randomized Controlled Trials (RCTs) are increasingly popular in the social sciences, not only in medicine. We believe. At best, an RCT yields an unbiased estimate, but this property is of limited practical value. Even then, estimates apply only to the sample selected for the trial, often no more than a convenience sample, and justification is required to extend the results to other groups, including any population to which the trial sample belongs, or to any individual including an individual in the trial. Demanding 'external validity' is unhelpful

At best, an RCT yields an unbiased estimate, but this property is of limited practical value. require minimal assumptions and can operate with little prior knowledge. This is an advantage when persuading distrustful audiences, but it is a disadvantage for cumulative scientific progress, where prior knowledge should be built upon, not discarded. RCTs can play a role in building scientific knowledge and useful predictions but they can only do so as part of a cumulative program, combining with other methods, including conceptual and theoretical development, to discover not 'what works', but 'why things work'.

Design	λ_{\max}	X	Root mean square error			
			A	B	C	D
Independent	1.03	1.00	1.00	1.00	1.00	1.00
Matched pairs	2.05	0.18	0.95	0.54	1.08	0.37
Deterministic	296.00	0.85	1.15	0.14	2.72	0.26
Rerand. 0.50	1.26	0.41	0.94	0.62	1.12	0.64
Rerand. 0.20	1.45	0.17	0.93	0.58	1.20	0.41
Rerand. 0.10	1.54	0.09	0.93	0.57	1.24	0.29
GS Walk 0.99	1.03	0.98	1.00	0.99	1.00	0.99
GS Walk 0.90	1.08	0.79	0.99	0.92	1.03	0.89
GS Walk 0.50	1.30	0.29	0.95	0.71	1.14	0.54
GS Walk 0.10	1.50	0.05	0.94	0.58	1.22	0.23
GS Walk 0.01	1.58	0.02	0.93	0.57	1.26	0.14

Regression adjustment by design

Theorem

For all potential outcomes, all covariates , and all robustness parameter values,

$$E[(\hat{\tau} - \tau)^2] \leq \frac{4L}{\phi n} \quad \text{where} \quad L = \min_{\beta \in \mathbb{R}^d} \left[\frac{1}{n} \|\mu - X\beta\|^2 + \frac{\xi^2 \phi}{(1-\phi)n} \|\beta\|^2 \right].$$

- L is the minimum of the loss function in a ridge regression of μ on X .
- The design implicitly controls for the covariates using ridge regression *before* running the experiment as if it knew all the potential outcomes, paying a cost of $1/\phi$.
- The parameter ϕ governs the trade-off between:
 - Linear fit (balance): $\|\mu - X\beta\|^2$.
 - Overfitting (robustness): $\|\beta\|^2$.

Confidence intervals

Theorem

Chernoff-type tail bound for the treatment effect estimator:

$$\Pr(|\hat{\tau} - \tau| \geq \gamma) \leq 2 \exp\left(\frac{-\gamma^2 n^2}{8L}\right) \quad \text{for all } \gamma > 0.$$

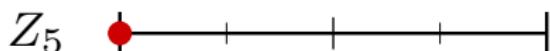
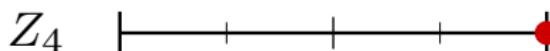
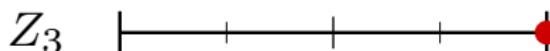
Corollary

Finite-sample valid $(1 - \alpha)$ -confidence interval:

$$\Pr(\hat{\tau} - \gamma_\alpha \leq \tau \leq \hat{\tau} + \gamma_\alpha) \geq 1 - \alpha,$$

where $\gamma_\alpha = \sqrt{8 \log(2/\alpha)L/n^2}$.

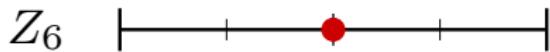
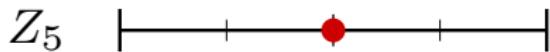
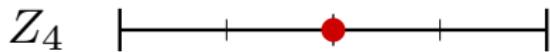
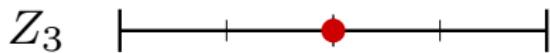
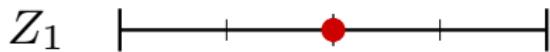
How does it work?



0

1

How does it work?

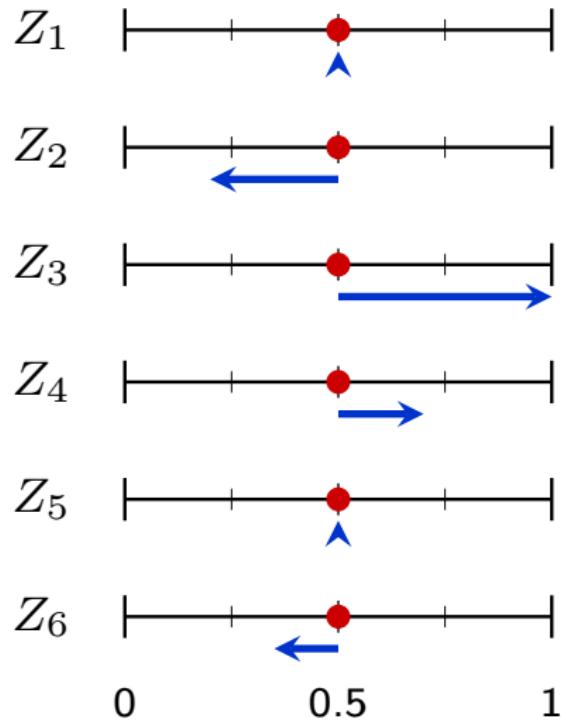


0

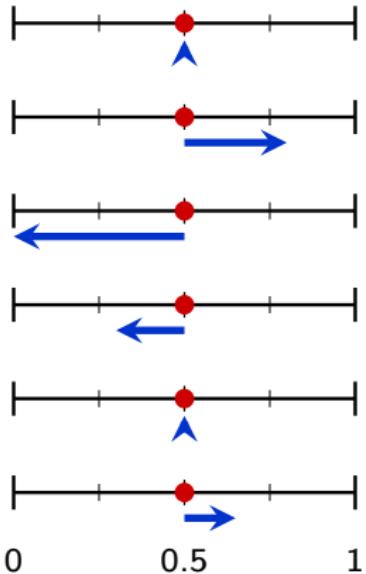
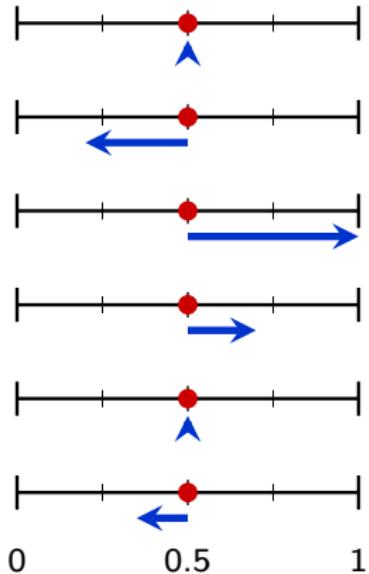
0.5

1

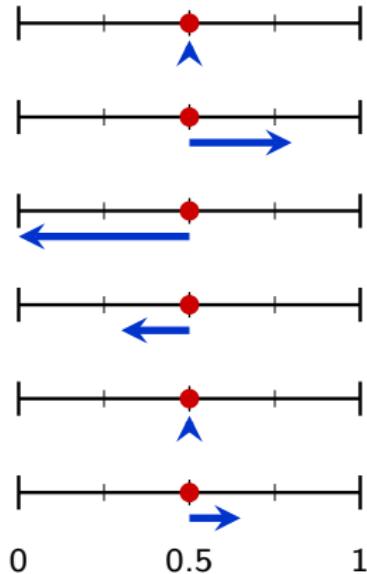
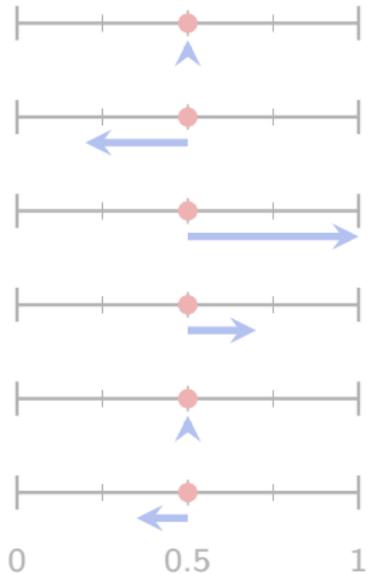
How does it work?



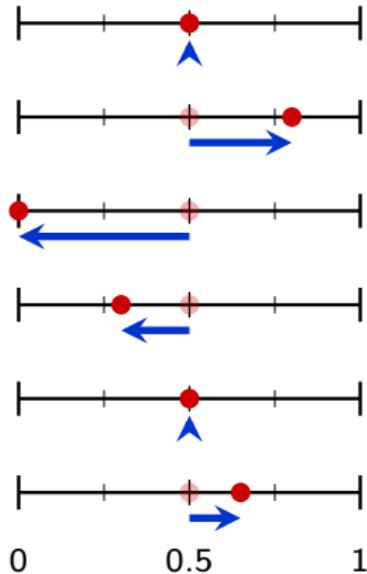
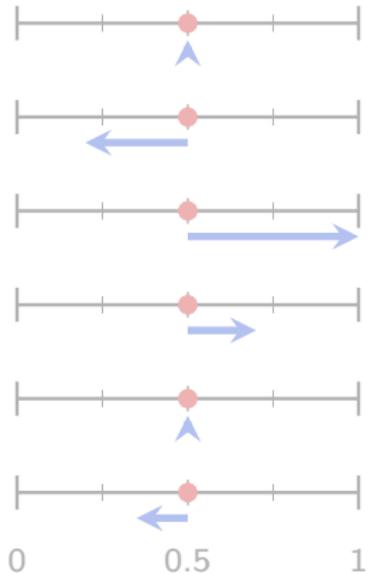
How does it work?



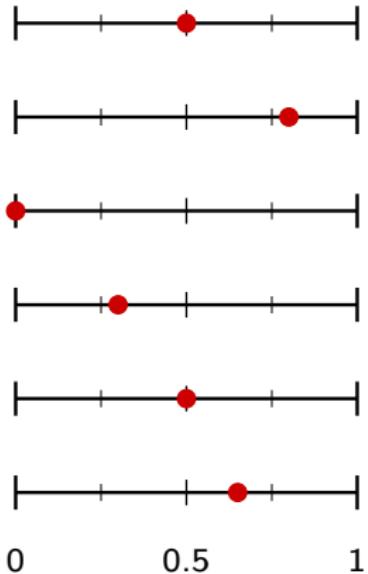
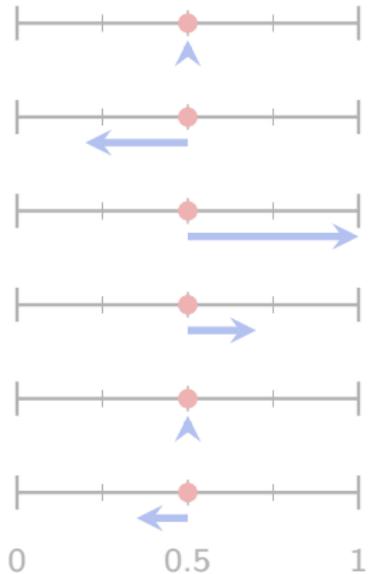
How does it work?



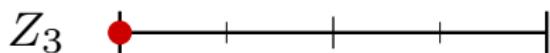
How does it work?



How does it work?



How does it work?



0

0.5

1