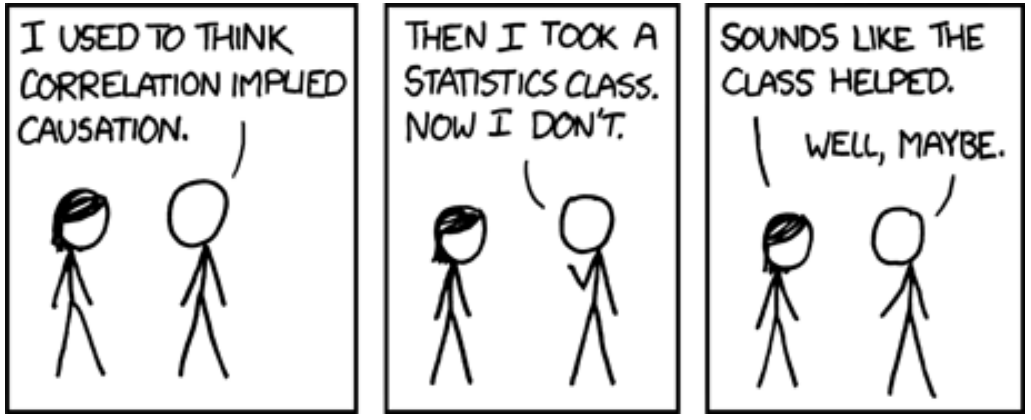
STA 235 - Causal Inference: Randomized Controlled Trials

Spring 2021

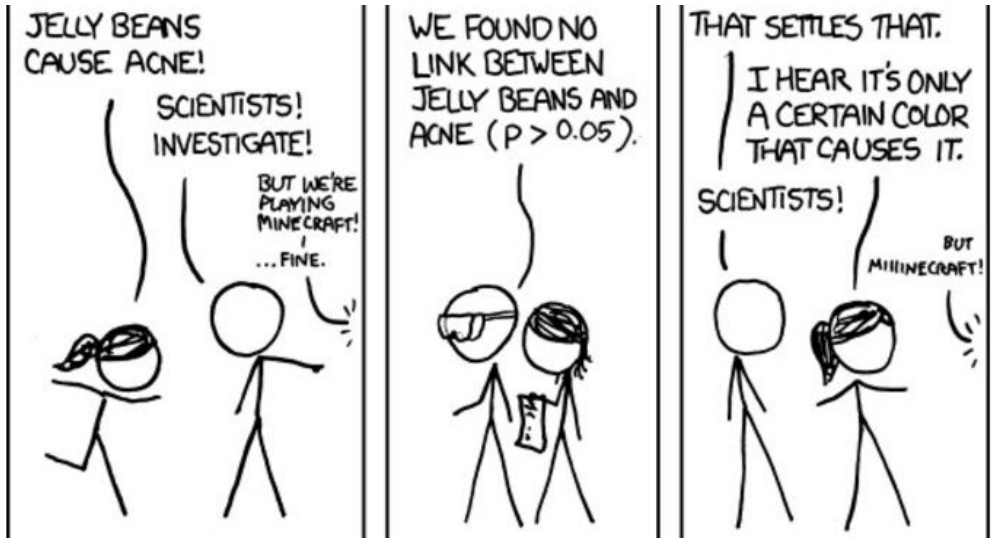
McCombs School of Business, UT Austin

Last week

- **Potential Outcomes Framework**
  - What are potential outcomes?
  - Definition of causal effects and estimands
- **Bias in Causal Inference**:
  - Ignorability assumption
  - Confounding vs Collider Bias



Today



- **Randomized Controlled Trials**:

Assumptions: The power of randomization

Design: What should we consider?

Limitations: Gold Standard?

The Magic of Randomization

The Fundamental Problem of Causal Inference

- Remember that we can only see one potential outcome
  - E.g. if Z is binary, either $Y(0)$ **OR** $Y(1)$

Fundamental Problem of Causal Inference

- Need for the **ignorability assumption**

$$Y(z) \perp\!\!\!\perp Z \quad \forall\, z \in Z$$

- Most times, **the ignorability assumption doesn't hold**
  - Why? (Remember the two types of bias we saw the previous class?)

Ignorability Assumption

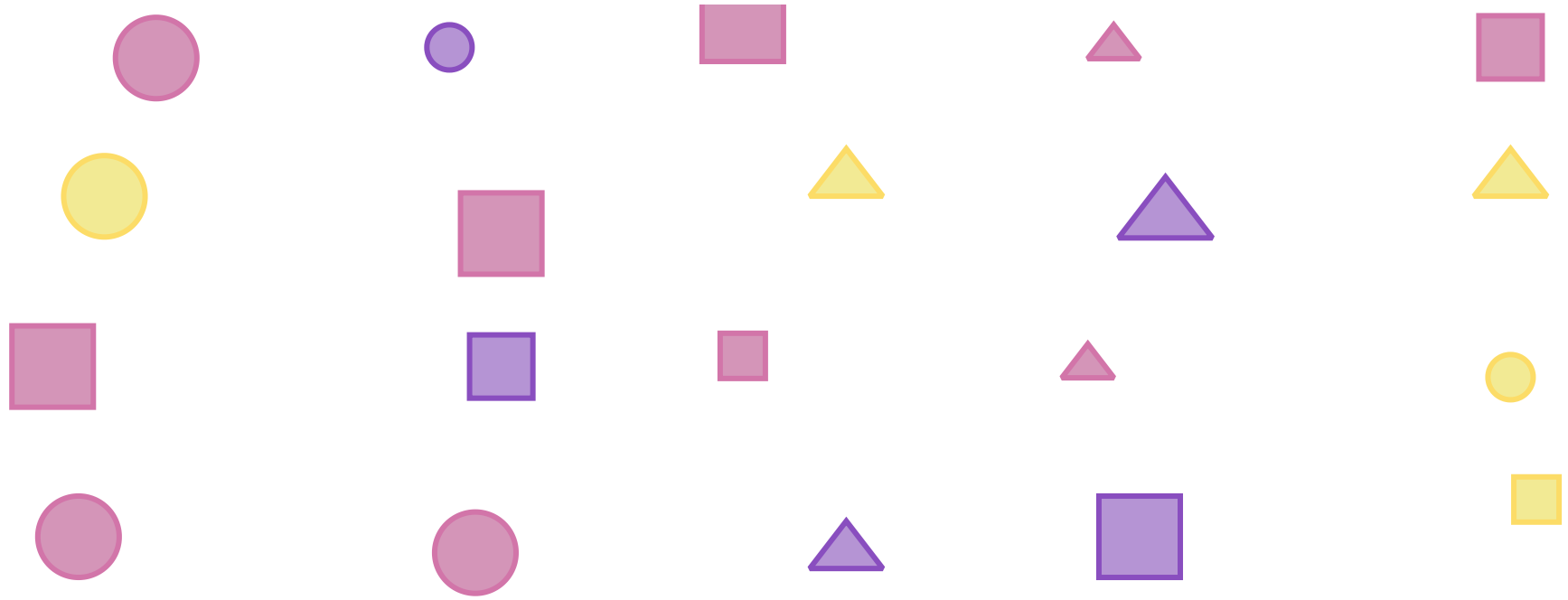
$$(Y(0), Y(1)) \perp\!\!\!\perp Z$$

- Most times, **the ignorability assumption doesn't hold**
  - For ATE:

Selection bias

Heterogeneous treatment effect bias

The problem with self-selection



Play

The power of randomization

- One way to make sure the ignorability assumption holds, is to do it by design:

Randomize the assignment of Z

i.e. Some units will **randomly** be chosen to be in the treatment group and others to be in the control group.

What does randomization buy us?

The power of randomization

- One way to make sure the ignorability assumption holds, is to do it by design:

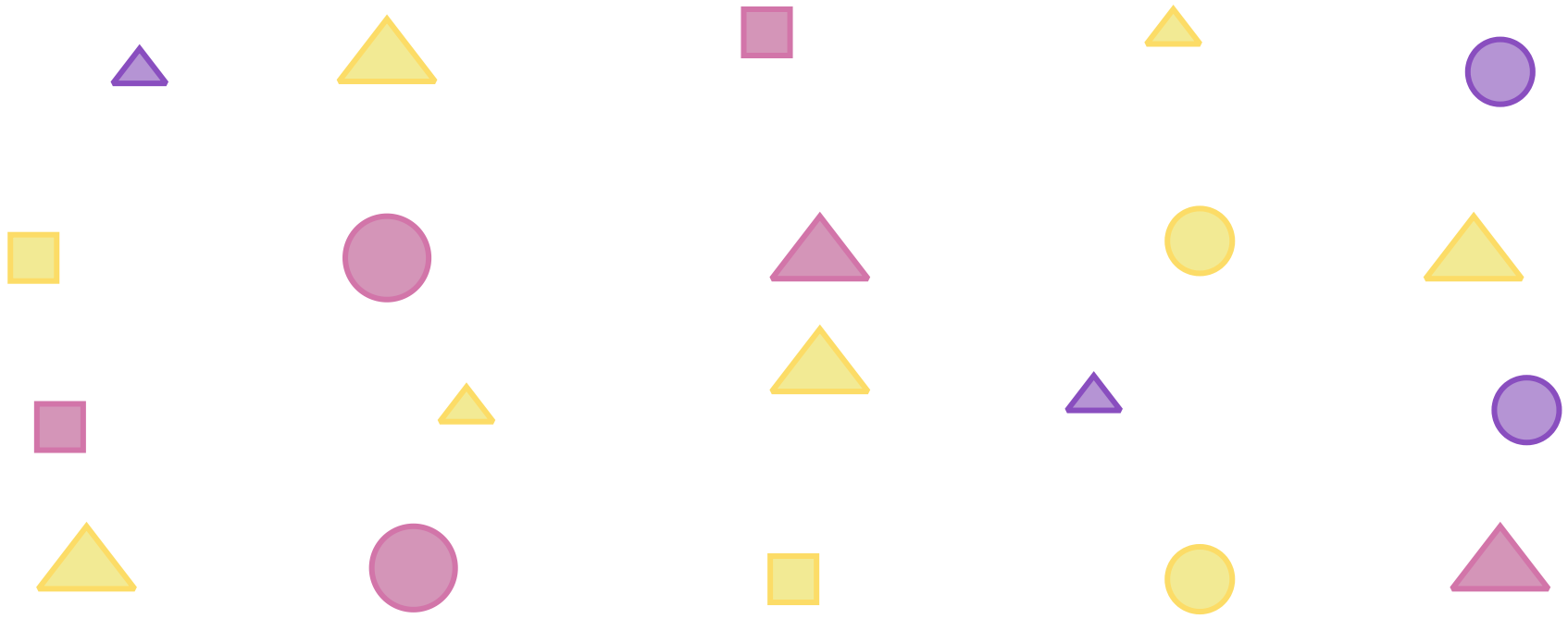
Randomize the assignment of Z

i.e. Some units will **randomly** be chosen to be in the treatment group and others to be in the control group.

What does randomization buy us?

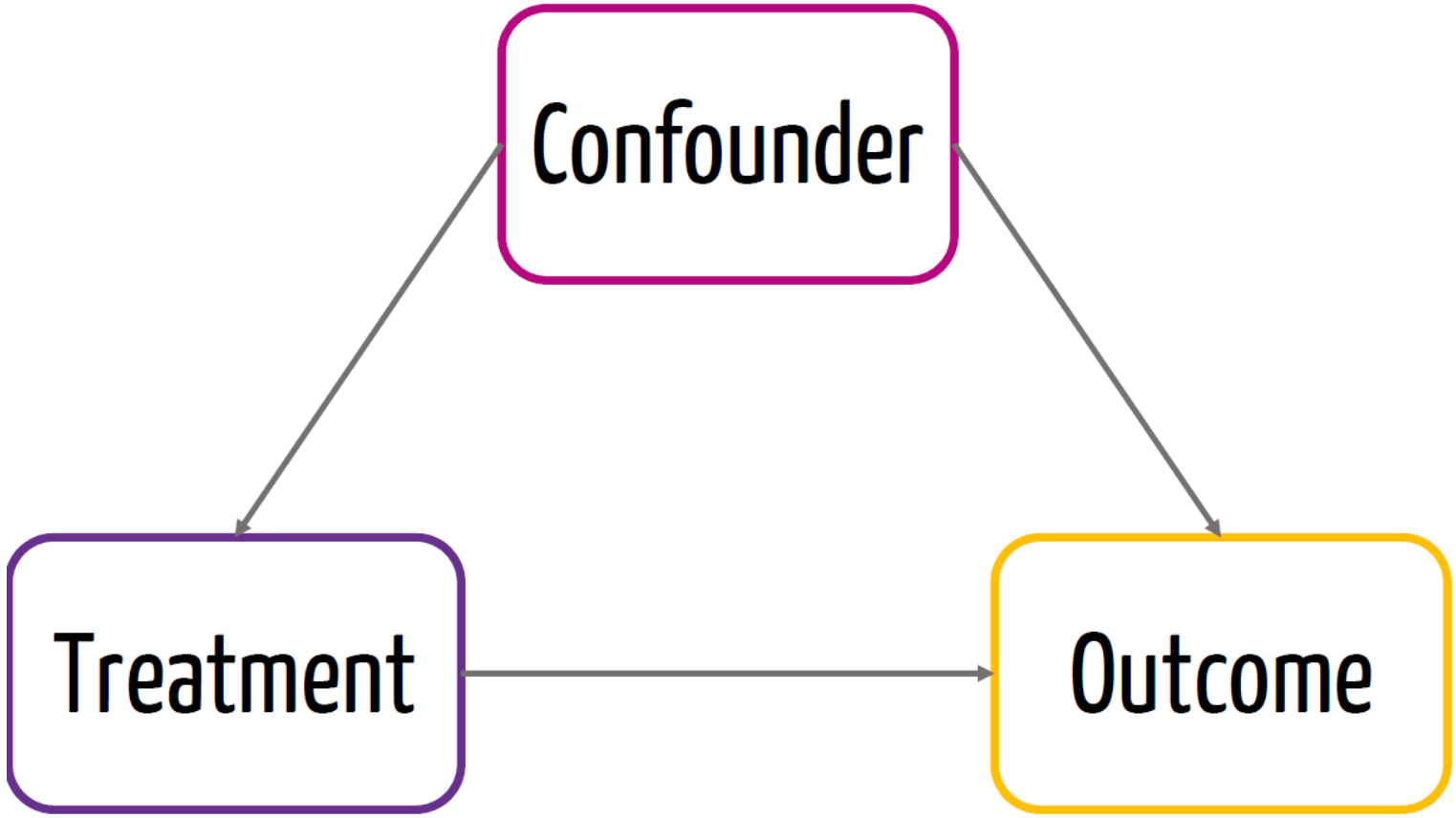
No (systematic) selection on observables OR unobservables

Randomization of z

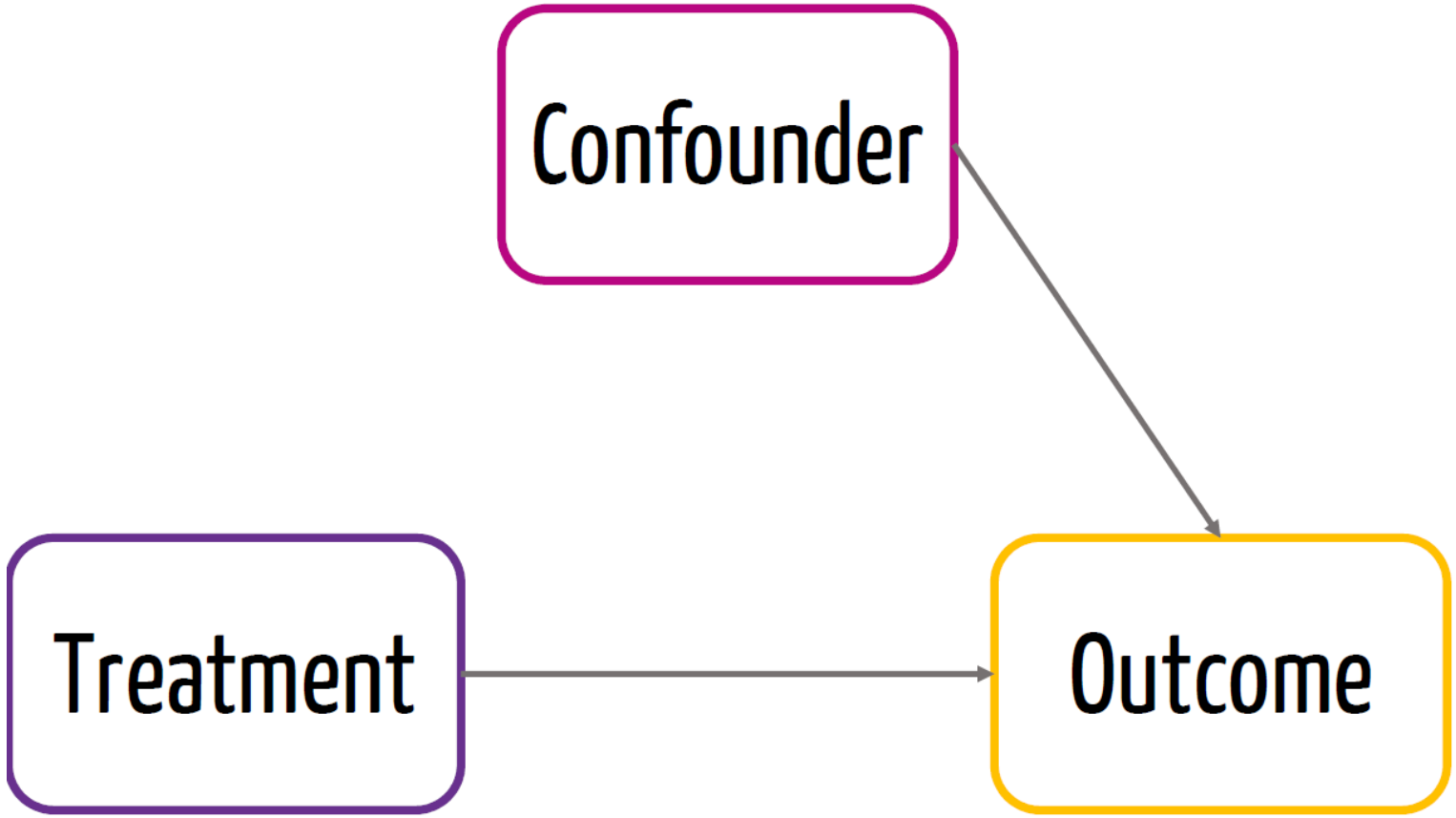


Play

Observational Causal Graph



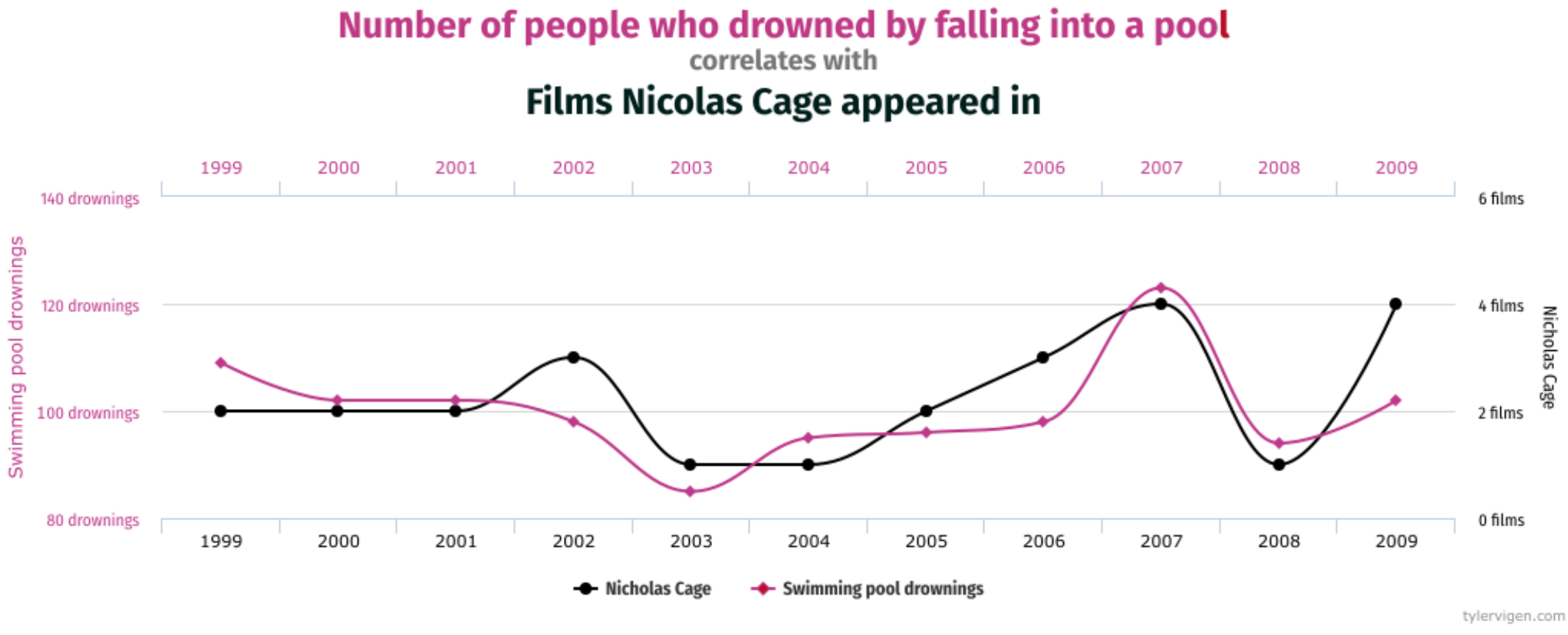
Experimental Causal Graph



If I randomize treatment allocation...

Can the treatment be potentially correlated with a confounder?

Just by chance!



RCTs: The Gold Standard

The New York Times

Nobel Economics Prize Goes to Pioneers in Reducing Poverty

Three professors, Abhijit Banerjee and Esther Duflo, both of M.I.T., and Michael Kremer of Harvard, were honored.

Abhijit Banerjee and Esther Duflo, both of M.I.T., and Michael Kremer of Harvard University won the Nobel Memorial Prize in Economic Sciences. Jonathan Nackstrand/Agence France-Presse — Getty Images

The Nobel went to economists who changed how we help the poor. But some critics oppose their big idea.

Randomized controlled trials and the debate over them, explained.

By Kelsey Piper | Dec 11, 2019, 9:00am EST

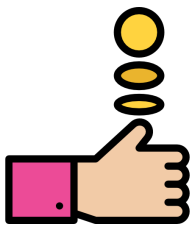
SHARE

The Laureates of The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel (L-R) Michael Kremer, Esther Duflo and Abhijit Banerjee pose after their Nobel Lectures at Stockholms University in Stockholm, Sweden, on December 8, 2019. | Photo by CHRISTINE OLSSON/TT News Agency/AFP via Getty Images

How do we randomize?

How do we randomize?

Coin flip*



How do we randomize?

Coin flip*

E.g., in R:

```
id <- seq(1,1000)

set.seed(100)

treat_id <- sample(id, length(id)*0.5)
control_id <- id[!(id %in% treat_id)]
```

How do we randomize?

Coin flip*

E.g., in R:

```
id <- seq(1,1000)

set.seed(100)

treat_id <- sample(id, length(id)*0.5)
control_id <- id[!(id %in% treat_id)]
```

How do we randomize?

Coin flip*

E.g., in R:

```
id <- seq(1,1000)

set.seed(100)

treat_id <- sample(id, length(id)*0.5)
control_id <- id[!(id %in% treat_id)]
```

How do we randomize?

Coin flip*

E.g., in R:

```
id <- seq(1,1000)

set.seed(100)

treat_id <- sample(id, length(id)*0.5)
control_id <- id[!(id %in% treat_id)]
```

How do we randomize?

Coin flip*

E.g., in R:

```
id <- seq(1,1000)

set.seed(100)

treat_id <- sample(id, length(id)*0.5)
control_id <- id[!(id %in% treat_id)]
```

```
z <- rep(0, length(id))
z[treat_id] <- 1

d <- data.frame("id" = id,"z" = z)

head(d)
```

```
##   id z
## 1  1 1
## 2  2 1
## 3  3 0
## 4  4 1
## 5  5 0
## 6  6 0
```

How would you randomize for more treatments?

How would you randomize for more treatments?

```r
# One way

id <- seq(1,1000)

set.seed(100)

types_treat <- c(0,1,2)

z <- sample(types_treat, length(id),
            replace = TRUE, prob = c(1/3,1/3,1/

table(z)
```

```
## z
##   0   1   2
## 354 306 340
```

How would you randomize for more treatments?

```r
# One way

id <- seq(1,1000)

set.seed(100)

types_treat <- c(0,1,2)

z <- sample(types_treat, length(id),
            replace = TRUE, prob = c(1/3,1/3,1/

table(z)
```

```
## z
##   0   1   2
## 354 306 340
```

```r
# Another way

id <- seq(1,1000)

set.seed(100)

rand <- runif(length(id))

id <- id[order(rand)]

z <- rep(c(0,1,2),1000/3 + 1)[1:length(id)]

table(z)
```

```
## z
##   0   1   2
## 334 333 333
```

How would you randomize for more treatments?

```r
# One way

id <- seq(1,1000)

set.seed(100)

types_treat <- c(0,1,2)

z <- sample(types_treat, length(id),
            replace = TRUE, prob = c(1/3,1/3,1/

table(z)
```

```
## z
##   0   1   2
## 354 306 340
```

```r
# Another way

id <- seq(1,1000)

set.seed(100)

rand <- runif(length(id))

id <- id[order(rand)]

z <- rep(c(0,1,2),1000/3 + 1)[1:length(id)]

table(z)
```

```
## z
##   0   1   2
## 334 333 333
```

How do we check that randomization was done correctly?

Checking for balance

```
library(designmatch)

d <- read.csv("https://raw.githubusercontent.com/maibennett/sta235/main/exampleSite/content/Classes/Week4/1_RCT/

head (round(d,3))
```

```
##   id z     x1     x2     x3     x4     x5     x6     x7     x8     x9    x10
## 1  1 1 -0.626 -0.897 -0.962  0.217 -0.841  0.270  2.287 -0.085 -0.767  0.019
## 2  2 1  0.184  0.185 -0.293 -0.542  1.384 -0.630 -1.197  0.840 -0.816 -0.184
## 3  3 0 -0.836  1.588  0.259  0.891 -1.255  0.869 -0.694 -0.463 -0.142 -1.371
## 4  4 1  1.595 -1.130 -1.152  0.596  0.070  1.727 -0.412 -0.551 -0.278 -0.599
## 5  5 0  0.330 -0.080  0.196  1.636  1.711  0.024 -0.971  0.736  0.436  0.295
## 6  6 0 -0.820  0.132  0.030  0.689 -0.603  0.368 -0.947 -0.108 -1.187  0.390
##      x11    x12    x13    x14    x15    x16    x17    x18    x19    x20
## 1 -0.591 -1.481  0.554 -0.662  0.259  0.476 -1.015  0.926 -1.189  1.163
## 2  0.027  1.577 -0.280  1.719  1.831 -0.125 -0.080  1.823  0.389 -0.586
## 3 -1.517 -0.957  1.775  2.122 -0.340  1.096 -0.233 -1.611 -0.344  1.785
## 4 -1.363 -0.920  0.187  1.497  0.897 -1.444 -0.817 -0.285 -0.548 -1.333
## 5  1.178 -1.998  1.143 -0.036  0.488  1.148  0.772 -0.342  0.981 -0.447
## 6 -0.934 -0.272  0.416  1.232 -1.255 -0.468 -0.166  0.366 -0.237  0.570
```

```
names_covs <- paste0("x",seq(1,20))
```

Checking for balance

```
meantab(d[,names_covs], d$z, which(d$z==1), which(d$
```

```
##     Mis   Min  Max Mean T Mean C Std Dif P-val
## x1    0 -3.01 3.81   0.04  -0.07    0.11  0.09
## x2    0 -2.72 3.01   0.06   0.06    0.00  0.98
## x3    0 -3.06 3.52   0.01   0.01    0.00  0.99
## x4    0 -2.84 3.17  -0.04  -0.03    0.00  0.96
## x5    0 -3.50 3.40   0.01   0.03   -0.02  0.77
## x6    0 -4.92 3.24   0.03  -0.08    0.12  0.07
## x7    0 -2.97 2.97  -0.03   0.03   -0.06  0.31
## x8    0 -3.28 2.98  -0.07  -0.02   -0.05  0.45
## x9    0 -3.04 2.76   0.01   0.00    0.02  0.77
## x10   0 -3.01 3.54   0.03  -0.01    0.04  0.55
## x11   0 -3.28 3.70  -0.02   0.04   -0.06  0.36
## x12   0 -3.05 3.11  -0.04  -0.01   -0.03  0.69
## x13   0 -2.84 3.59  -0.01   0.01   -0.02  0.71
## x14   0 -2.96 2.92  -0.02  -0.04    0.01  0.82
## x15   0 -4.21 3.44   0.04   0.04    0.00  0.98
## x16   0 -3.33 3.47   0.04   0.04    0.00  0.95
## x17   0 -3.56 3.94   0.05  -0.01    0.06  0.33
## x18   0 -3.74 3.17  -0.08   0.00   -0.08  0.20
## x19   0 -3.03 2.87   0.02  -0.03    0.04  0.49
## x20   0 -3.00 2.93  -0.03   0.00   -0.04  0.58
```

- **Columns of interest**:
  - Mean T: Avg. for treat. group
  - Mean C: Avg. for control group
  - Std. Dif: Difference T-C (in SD)
  - P-val: Whether diff is significant

Checking for balance

```
meantab(d[,names_covs], d$z, which(d$z==1), which(d$
```

```
##     Mis   Min  Max Mean T Mean C Std Dif P-val
## x1    0 -3.01 3.81   0.04  -0.07    0.11  0.09
## x2    0 -2.72 3.01   0.06   0.06    0.00  0.98
## x3    0 -3.06 3.52   0.01   0.01    0.00  0.99
## x4    0 -2.84 3.17  -0.04  -0.03    0.00  0.96
## x5    0 -3.50 3.40   0.01   0.03   -0.02  0.77
## x6    0 -4.92 3.24   0.03  -0.08    0.12  0.07
## x7    0 -2.97 2.97  -0.03   0.03   -0.06  0.31
## x8    0 -3.28 2.98  -0.07  -0.02   -0.05  0.45
## x9    0 -3.04 2.76   0.01   0.00    0.02  0.77
## x10   0 -3.01 3.54   0.03  -0.01    0.04  0.55
## x11   0 -3.28 3.70  -0.02   0.04   -0.06  0.36
## x12   0 -3.05 3.11  -0.04  -0.01   -0.03  0.69
## x13   0 -2.84 3.59  -0.01   0.01   -0.02  0.71
## x14   0 -2.96 2.92  -0.02  -0.04    0.01  0.82
## x15   0 -4.21 3.44   0.04   0.04    0.00  0.98
## x16   0 -3.33 3.47   0.04   0.04    0.00  0.95
## x17   0 -3.56 3.94   0.05  -0.01    0.06  0.33
## x18   0 -3.74 3.17  -0.08   0.00   -0.08  0.20
## x19   0 -3.03 2.87   0.02  -0.03    0.04  0.49
## x20   0 -3.00 2.93  -0.03   0.00   -0.04  0.58
```

Did the randomization work?

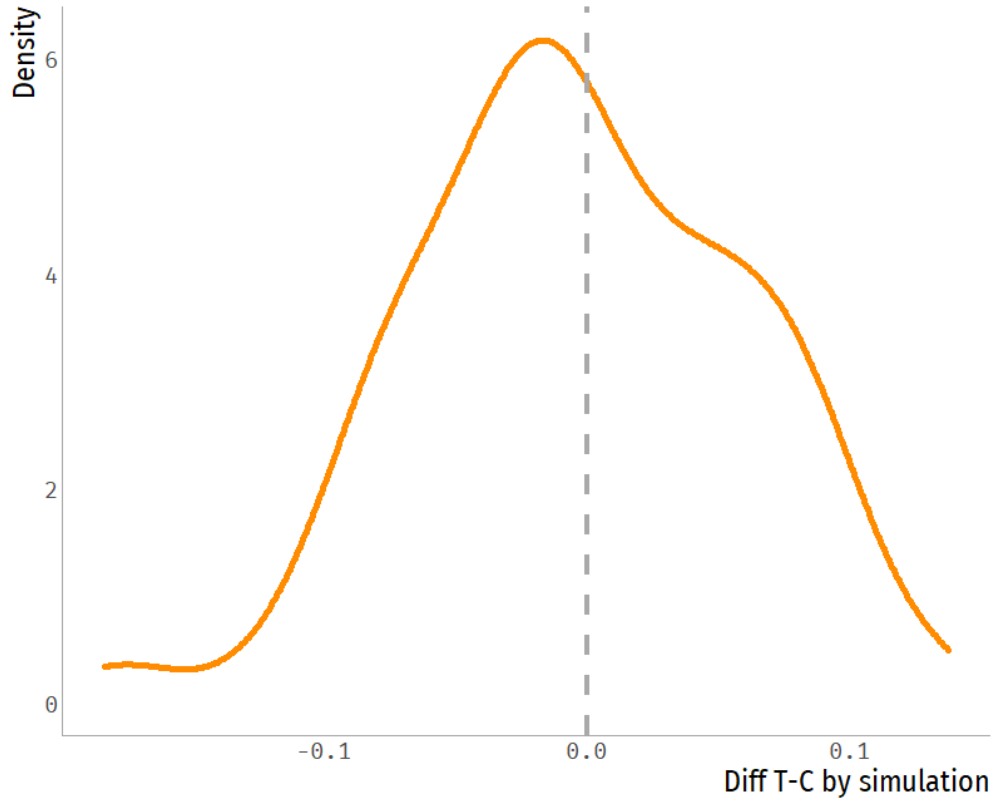
Why are there significant differences?

Randomization assures balance *in expectation*

Let's go to R

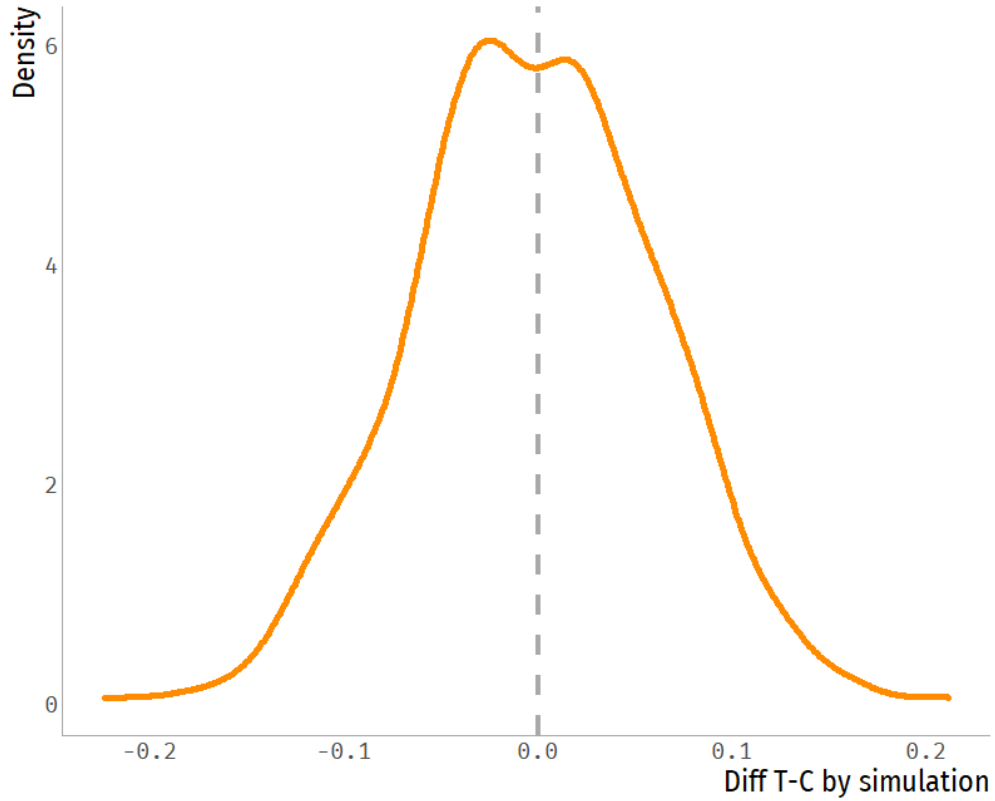
Simulations: Difference between T and C for X1

- 100 random assignments for Z



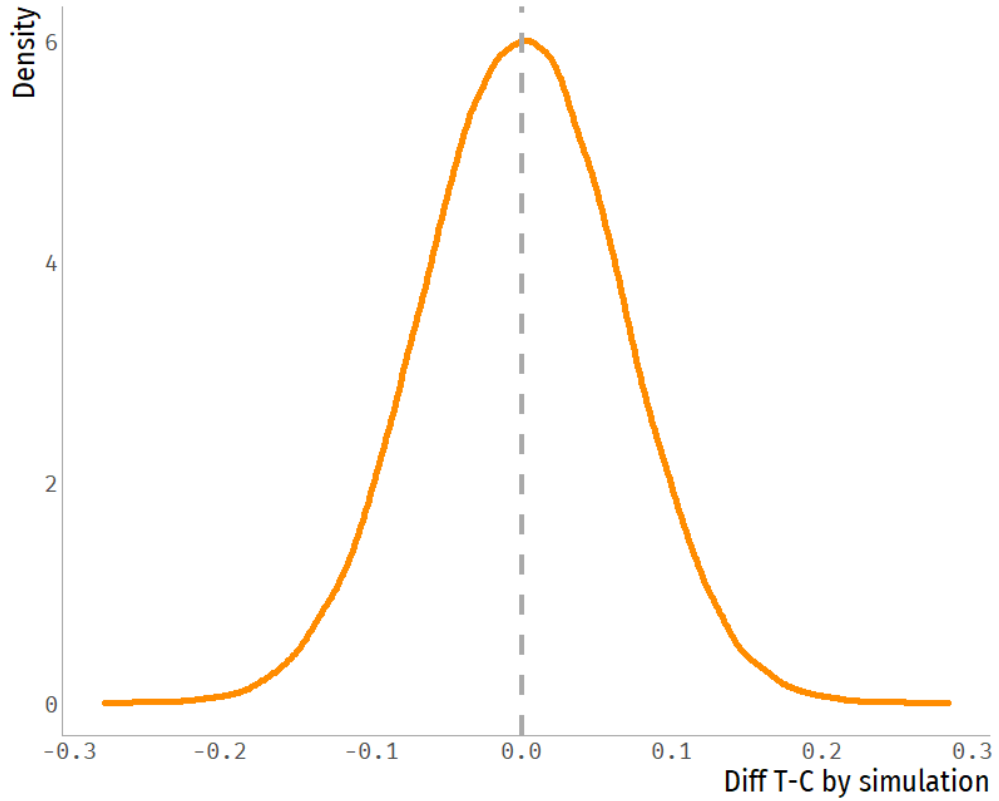
Simulations: Difference between T and C for X1

- 1,000 random assignments for Z



Simulations: Difference between T and C for X1

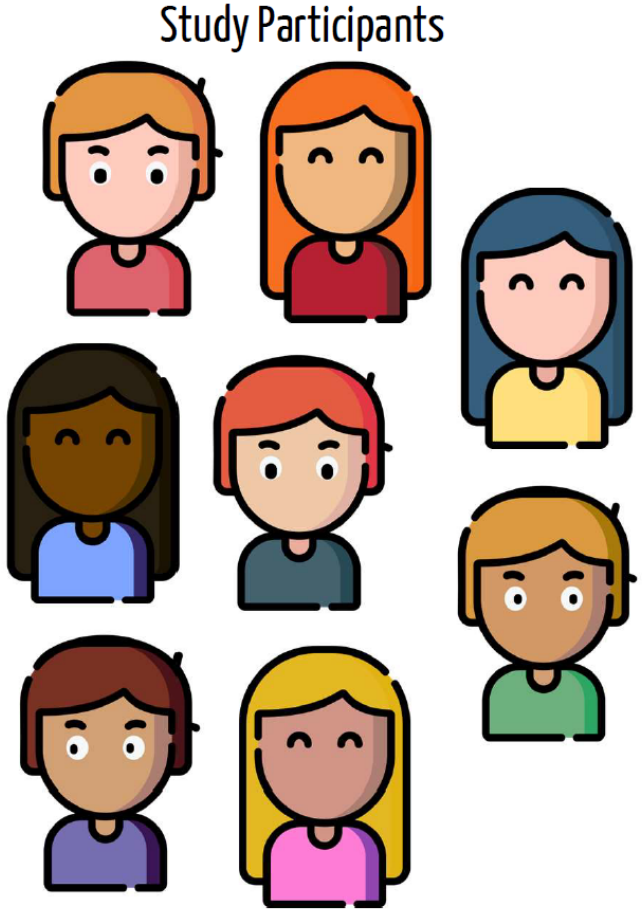
- 100,000 random assignments for Z



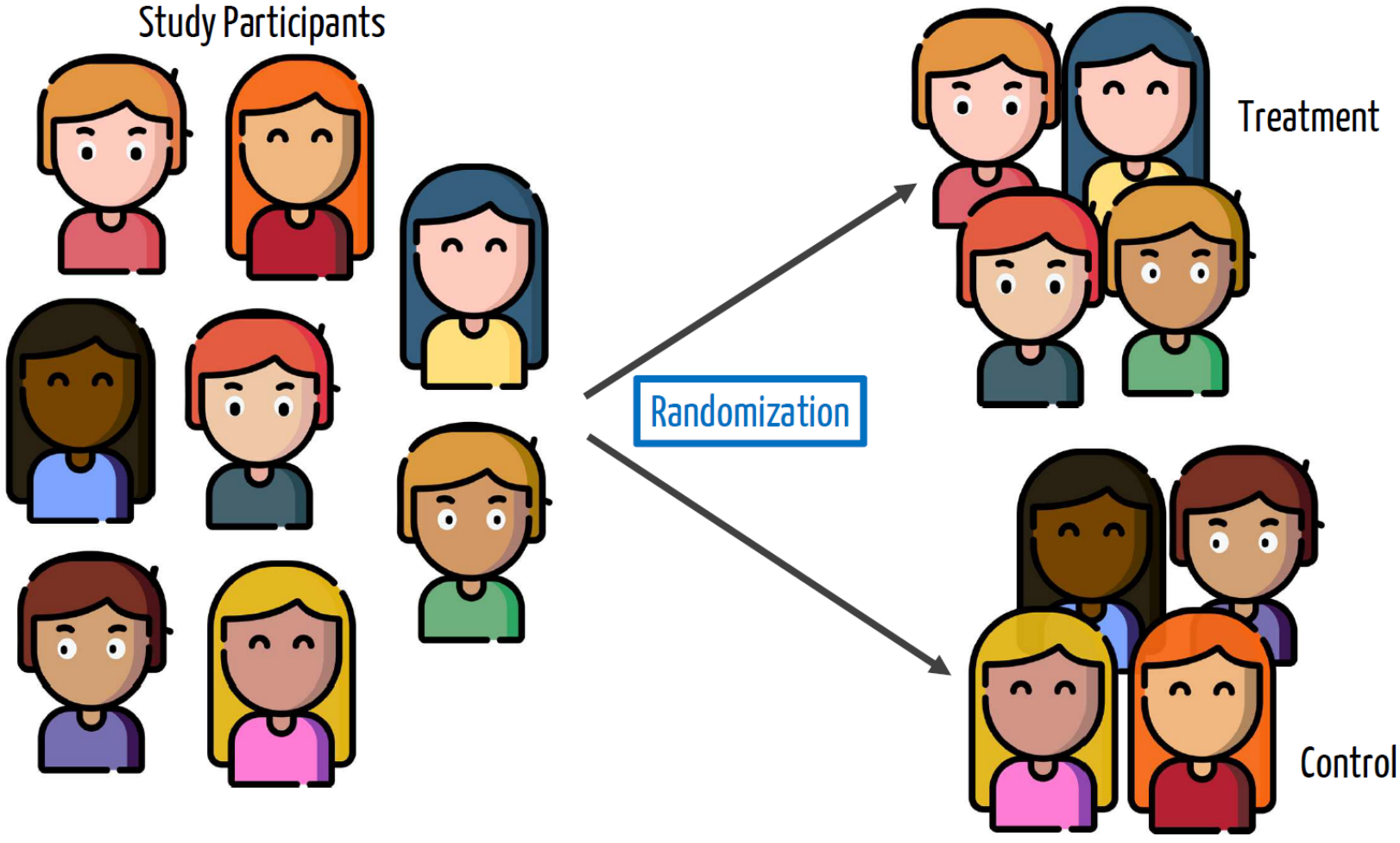
Can we ensure balance?

- In RCTs, in general, you can't ensure balance on all covariates.
- But you can **stratify**!
- Stratification means dividing your data into different stratas or groups *S*, based on one or more covariates.
- Then, you randomize in the same way *within strata*.

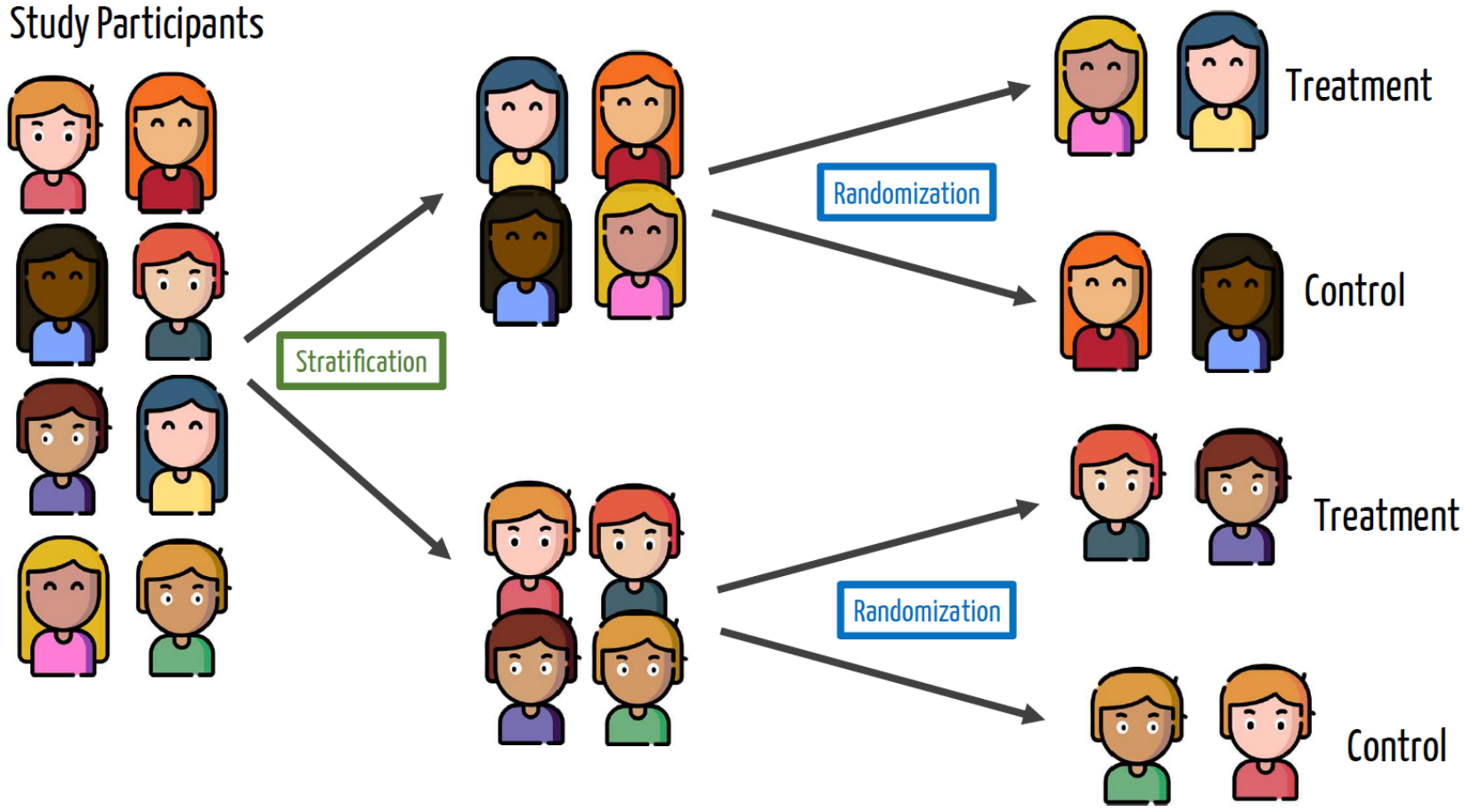
Randomization of treatment



Randomization of treatment with no strata



Randomization of treatment with strata



RCTs in Practice

How to analyze RCTs?

How to analyze RCTs?

Easy! (Statistically speaking)

How to analyze RCTs?

Easy! (Statistically speaking)

1) Check for balance

How to analyze RCTs?

Easy! (Statistically speaking)

1) Check for balance

2) Calculate difference in sample means between treatment and control group

Let's look at some data

- "Get out the Vote" Large-Scale Mobilization experiment (Arceneaux, Gerber, and Green, 2006)
    - "Households containing one or two registered voters where **randomly assigned** to treatment or control groups"
    - Treatment: GOTV phone calls
    - Stratified RCT: Two states divided into competitive and noncompetitive

Checking for balance

Let's go to R

Checking for balance

Balance Table by Stratum

	Non-competitive		Competitive		Non-competitive		Competitive	
	Treat	**Control**	**Treat**	**Control**	**Treat**	**Control**	**Treat**	**Control**
female2	0.552	0.546	0.541	0.535	0.549	0.545	0.543	0.541
fem_miss	0	0	0	0	0.026	0.025	0.022	0.021
age	52.157	51.977	50.81	50.862	55.795	55.782	53.481	53.464
newreg	0.117	0.116	0.133	0.134	0.048	0.049	0.048	0.046
persons	1.496	1.497	1.513	1.518	1.539	1.538	1.529	1.533
contact	0	0.369	0	0.388	0	0.461	0	0.452
vote98	0.231	0.227	0.258	0.259	0.572	0.574	0.599	0.594
vote00	0.564	0.567	0.595	0.593	0.734	0.732	0.781	0.78

Estimating the effect

- Depending on the design, usually you can **compare group means** or fit a **simple regression**

$$\frac{1}{N_T}\sum_{i \in T} Y_i - \frac{1}{N_C}\sum_{i \in C} Y_i$$

$$Y_i = \beta_0 + \beta_1 Z_i + \varepsilon_i$$

How do we incorporate stratification here?

Estimating the effect

- In stratified RCTs, we need to consider the strata!
- Run a regression with *fixed effects by strata*

$$Y_i = \beta_0 + \beta_1 Z_i + \gamma_s + \varepsilon_i$$

Estimating the effect

```
library(estimatr)

d_s1$strata <- interaction(d_s1$state, d_s1$competiv)

summary(estimatr::lm_robust(vote02 ~ treat_real + strata, data = d_s1))
```

```
##
## Call:
## estimatr::lm_robust(formula = vote02 ~ treat_real + strata, data = d_s1)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error   t value Pr(>|t|) CI Lower CI Upper      DF
## (Intercept) 0.509915  0.0004633 1100.6000   0.0000  0.50901 0.510823 1905315
## treat_real  0.001499  0.0020709    0.7237   0.4693 -0.00256 0.005557 1905315
## strata1.1   0.085625  0.0016372   52.2989   0.0000  0.08242 0.088834 1905315
## strata0.2   0.049105  0.0009801   50.1034   0.0000  0.04718 0.051026 1905315
## strata1.2   0.146435  0.0009803  149.3733   0.0000  0.14451 0.148357 1905315
##
## Multiple R-squared:  0.01172 ,    Adjusted R-squared:  0.01172
## F-statistic:  5937 on 4 and 1905315 DF,  p-value: < 2.2e-16
```

Estimating the effect

- One important thing to note in the previous analysis is that **assignment to treatment** $\neq$ **contact**

```
table(d_s1$treat_real, d_s1$contact)
```

```
##
##             0       1
##   0 1845348       0
##   1   34929   25043
```

Does this affect the internal validity of the study?

When we assume...

Other assumptions

- We already talked about the **ignorability assumption**
- **Stable Unit Treatment Value Assumption (SUTVA)**:
  - No interference
  - No hidden variations of treatments

SUTVA: No interference

- *"The treatment applied to one unit does not affect the outcome for other units"*
- Imagine we have two individuals, 1 and 2:
  - Z_1 and Z_2 will be the treatment assignment for 1 and 2, respectively.
  - Under SUTVA:

$$(Y_2(0), Y_2(1)) \perp Z_1$$

SUTVA: No hidden variations of treatments

- *"An individual receiving an specific treatment level cannot receive different forms of that treatment."*
- Think about the headache example from last class:
  - Individual 1 gets a **new aspirin (aspirin +)**
  - Individual 2 gets an **old aspirin (aspirin -)**
  - Individual 3 doesn't get a pill
- If we label the treatments as $Z = 0$ (doesn't get an aspirin) and $Z = 1$ (gets an aspirin), **potential outcomes would change depending on whether they got** *aspirin +* **or** *aspirin -*.

When SUTVA hits the fan

When do you think SUTVA could fail?

h/t to @paul_gp

When SUTVA hits the fan

Network effects

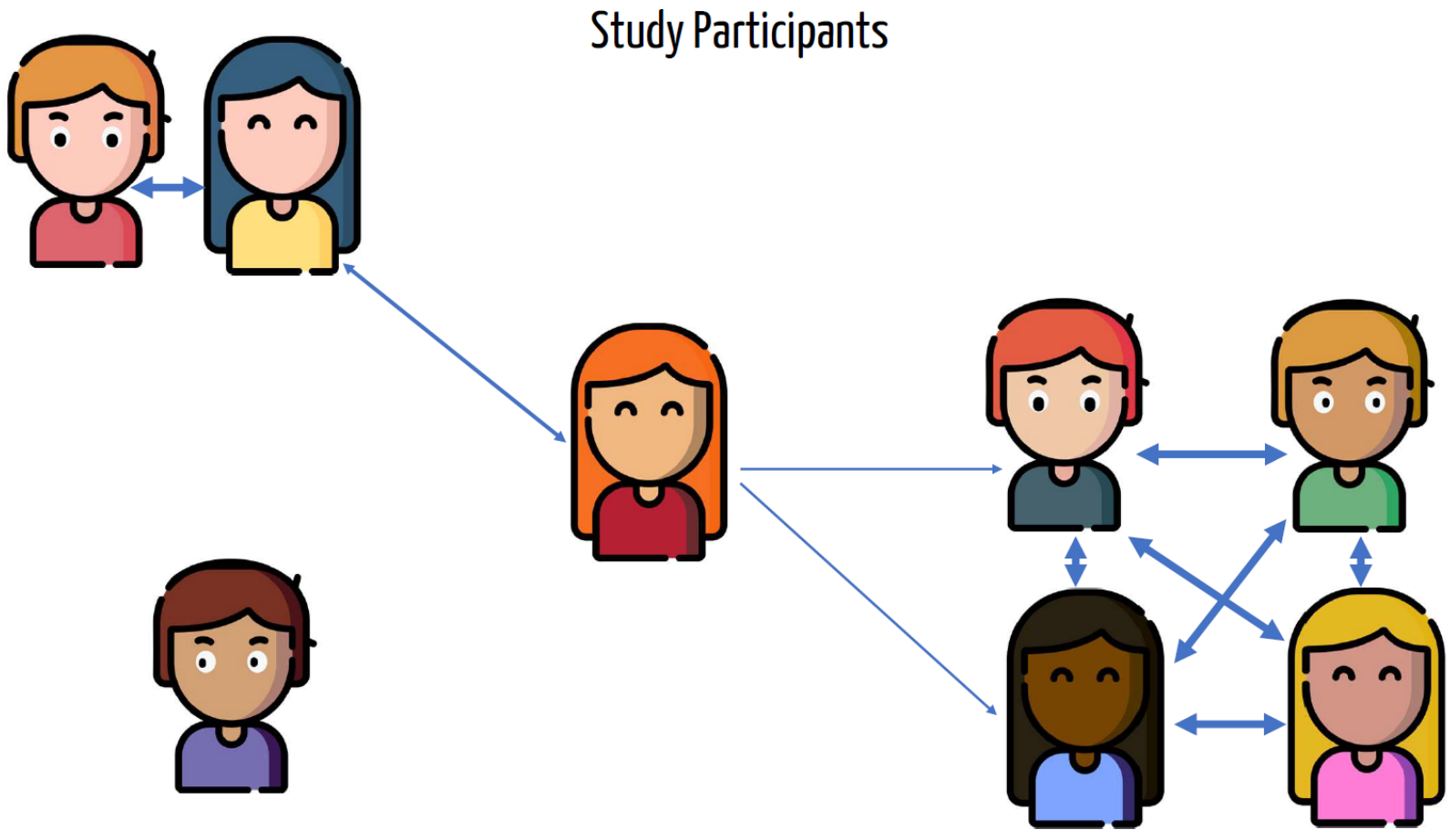
General Equilibrium Effects

Network effects

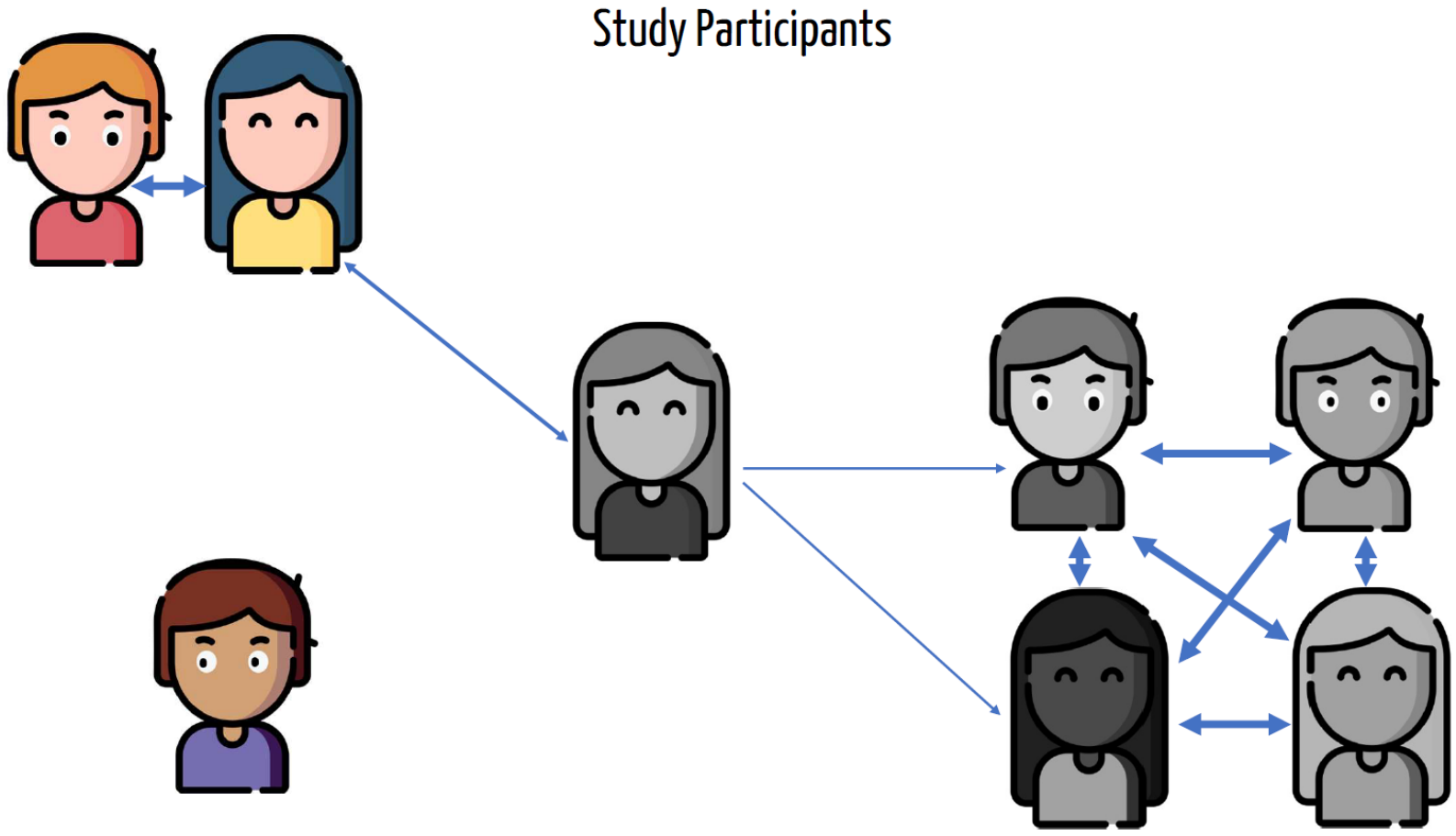
- Also referred to as **spillovers**
- Potential outcomes will depend on *who gets the treatment*

Let's look at an example

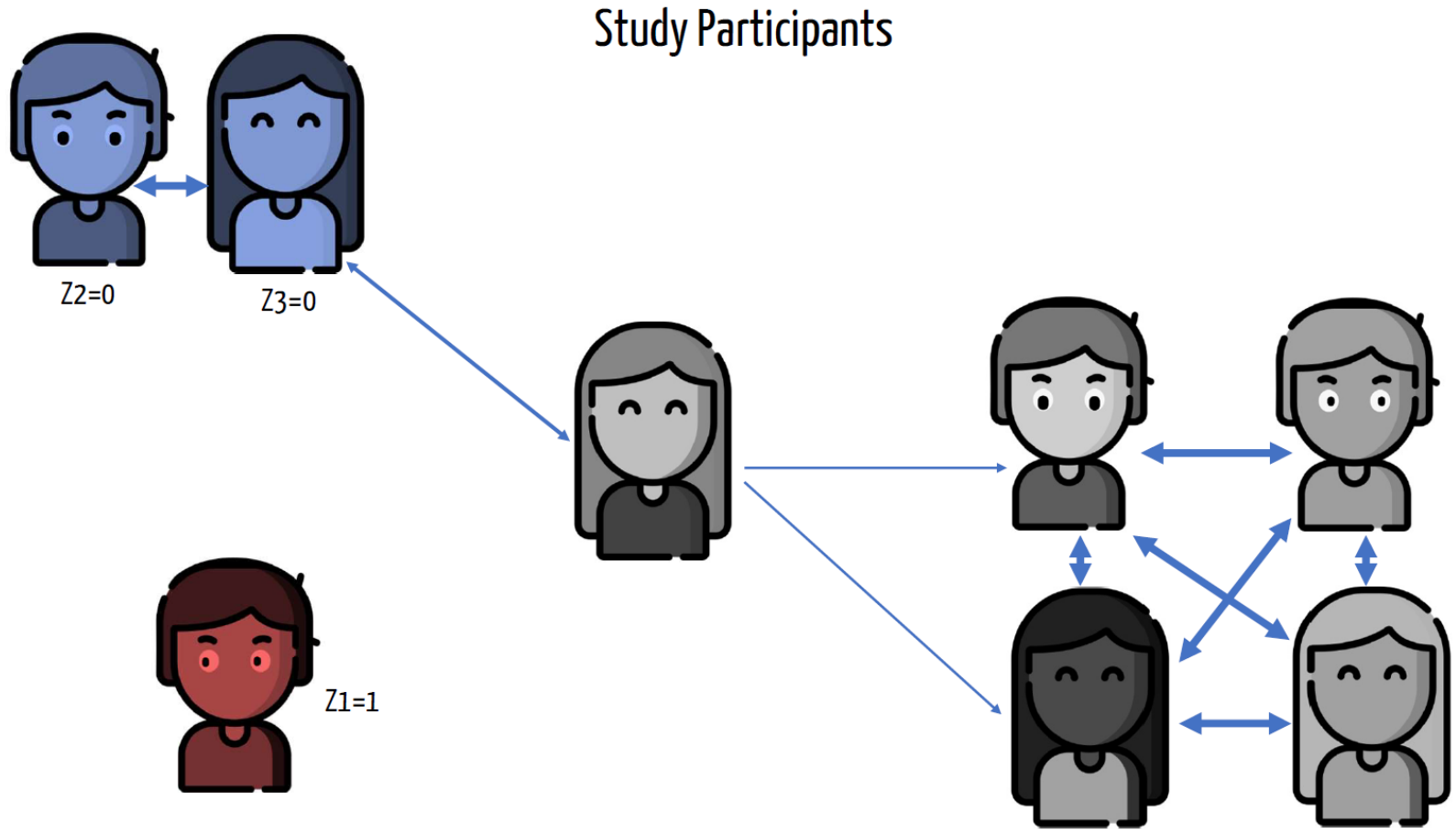
Network effects



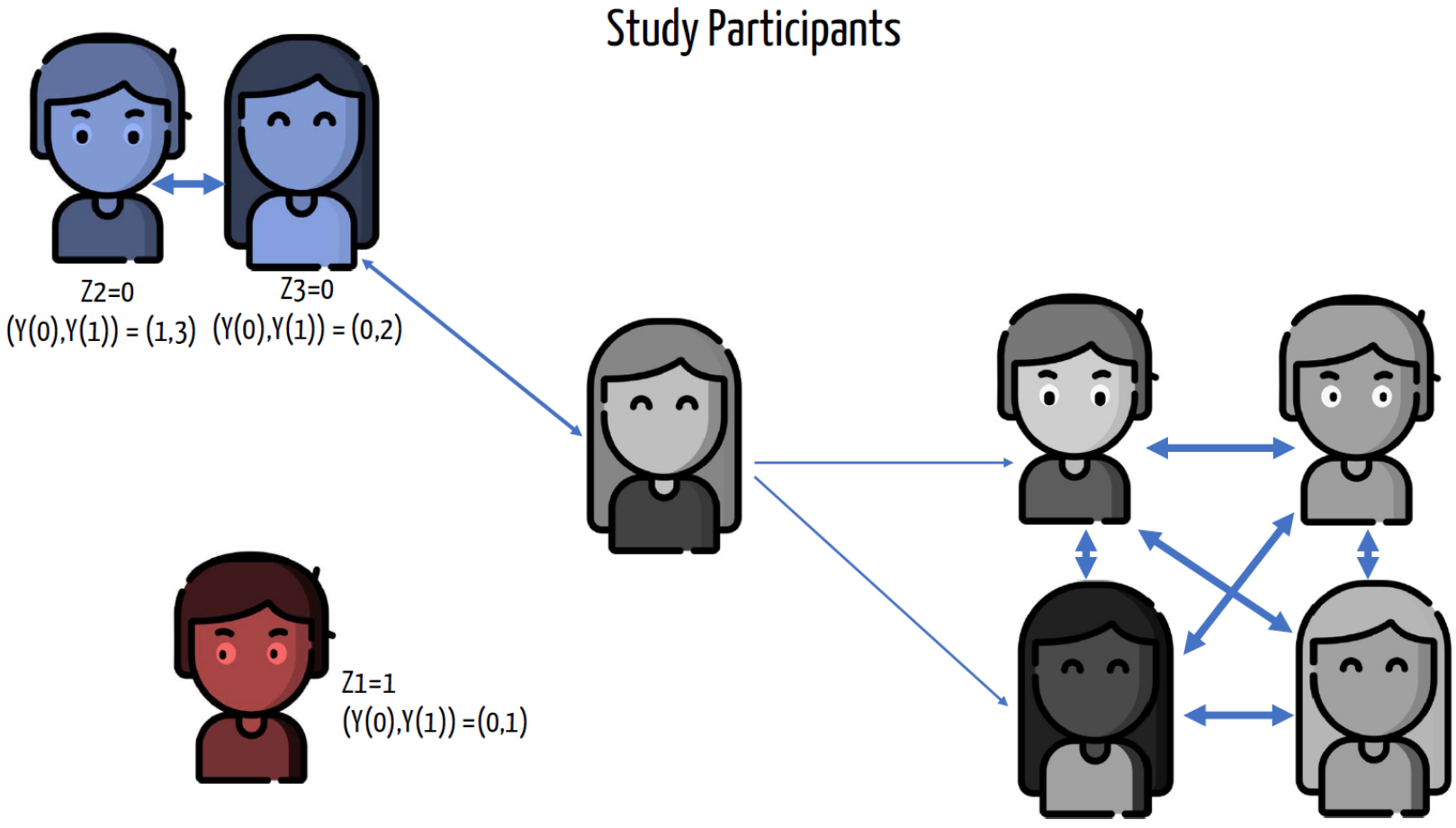
Network effects



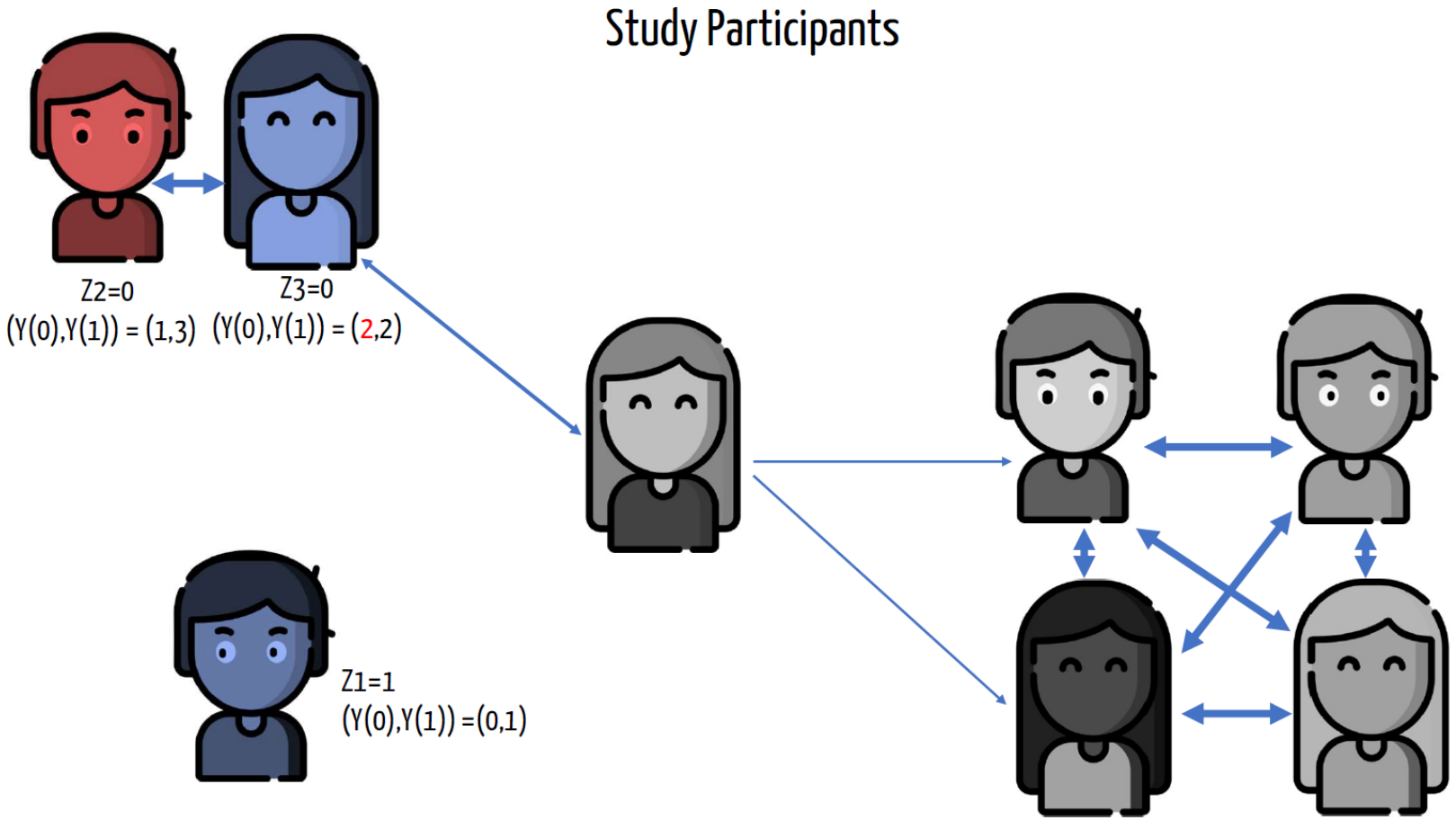
Network effects



Network effects



Network effects



Network effects

Can we do something about this?

1. **Randomize at a higher level** (e.g. neighborhood, school, etc. instead of at the individual level)
    - Note that you will have to **cluster your standard errors**
2. **Model the network!**

General Equilibrium Effects

- Usually arise when you **scale up** a program or intervention.
- Imagine you want to test the effect of providing information about employment and expected income to students to see whether it affect their choice of university and/or major.

What could happen if you offer it to everyone?

Show me the power

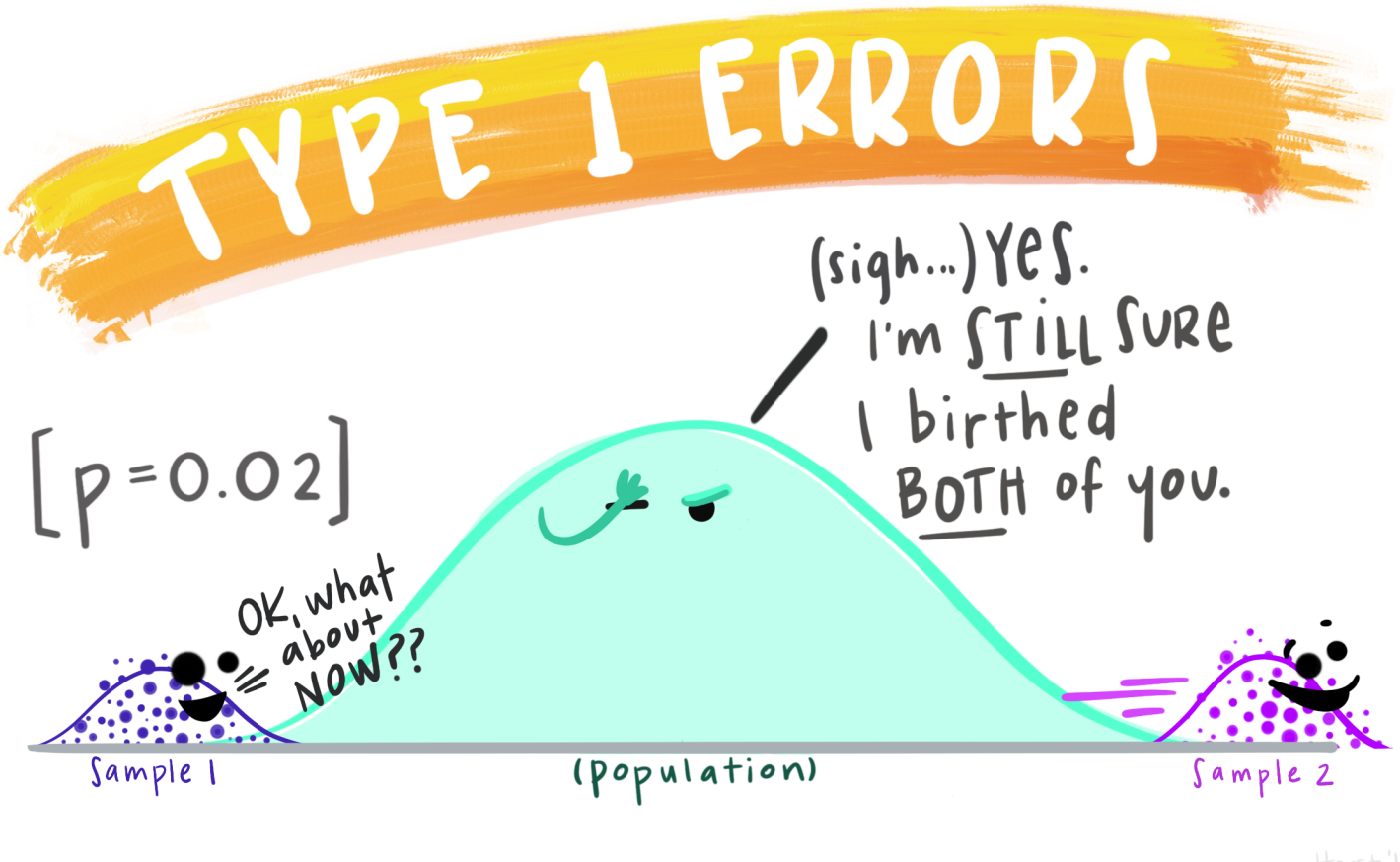
Statistical power

- **Statistical power** refers to the probability that a test correctly rejects the null hypothesis H_0, when H_0 is false.
    - It's related to sample size!

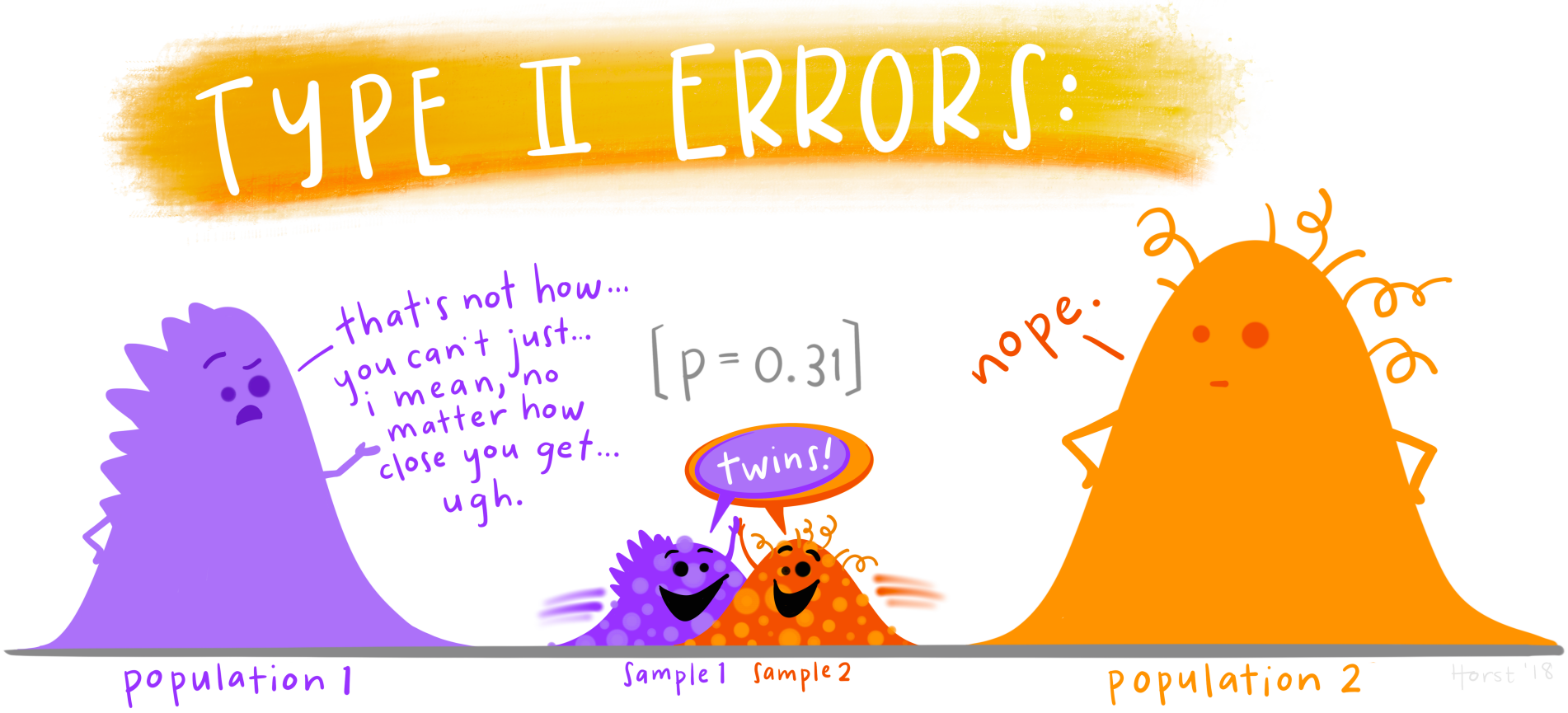
Power = 1- Pr(Type II error)

- **Type I error**: Probability of rejecting the null hypothesis, when the null is true (α).
- **Type II error**: Probability of not rejecting the null hypothesis, when the null is false (β).

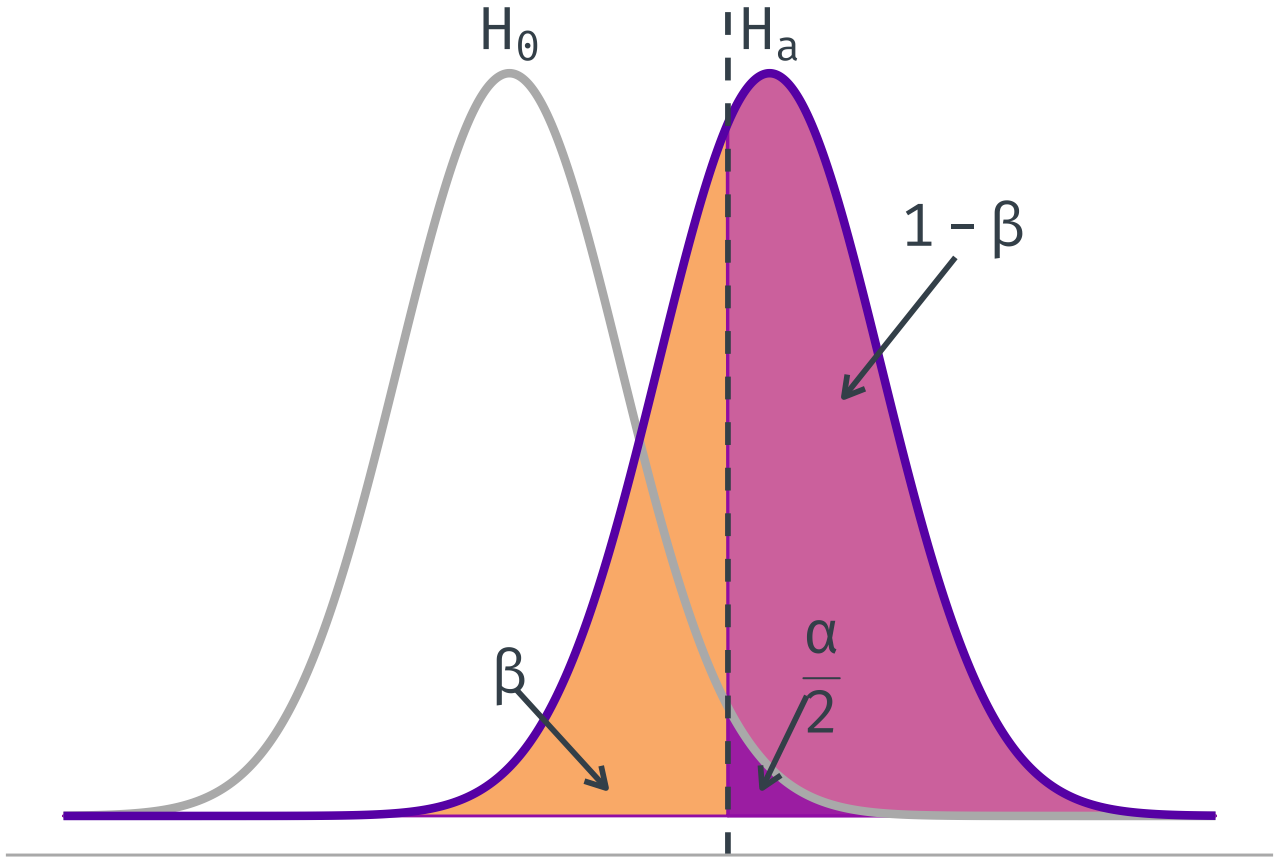
Statistical power



Statistical power



Statistical power in pictures



Statistical power and sample size

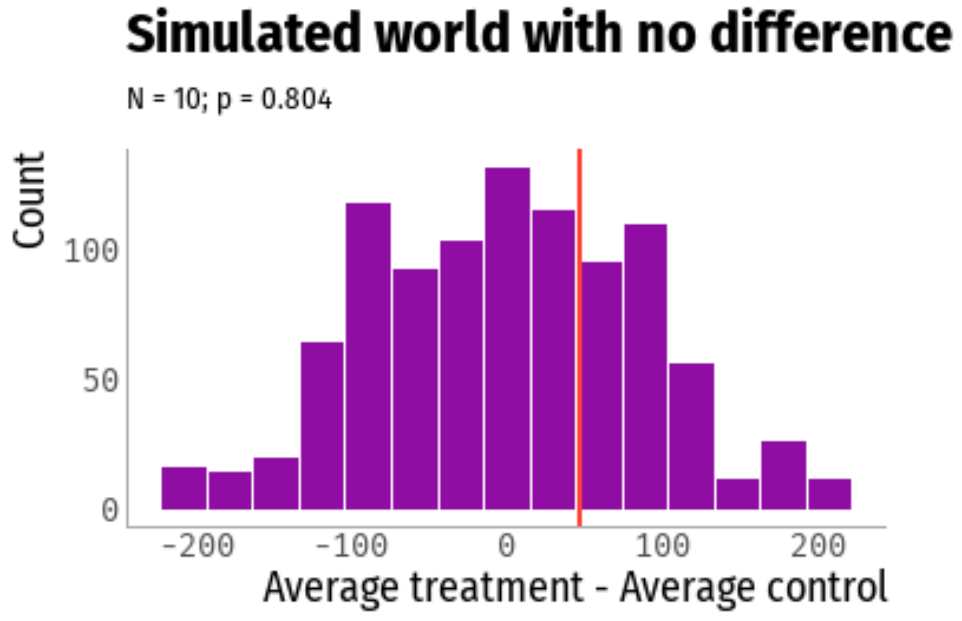
How big of a sample?

A training program causes incomes to rise by $40

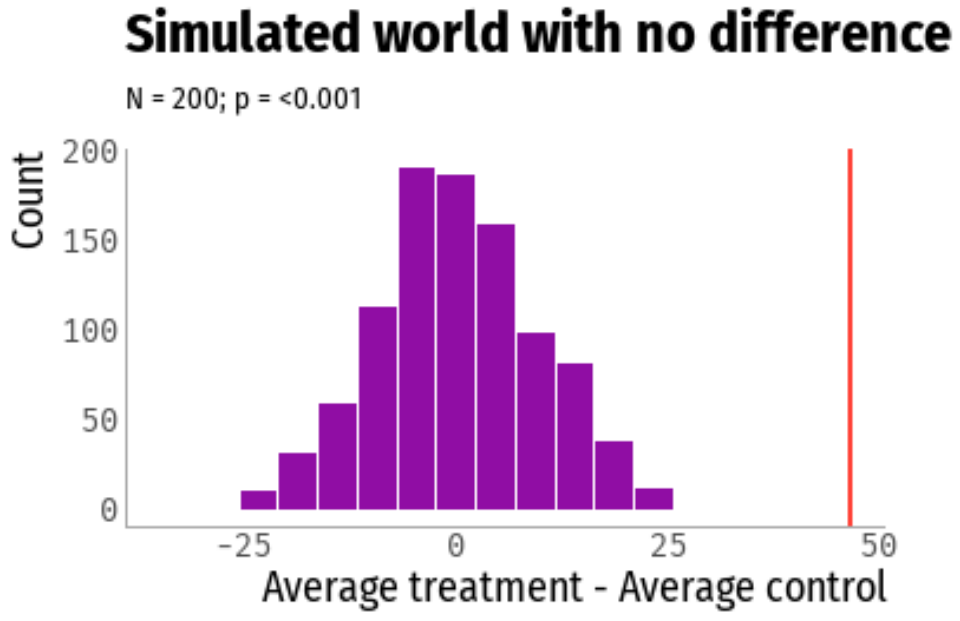
Person	Group	Before	After	Difference
295	Control	122.09	219.04	96.95
126	Treatment	205.60	199.84	-5.76
400	Control	133.25	120.40	-12.85
94	Treatment	270.11	206.56	-63.54
250	Control	344.37	212.89	-131.49
59	Treatment	312.41	268.06	-44.35

Can I detect an effect?

Enroll 10 participants

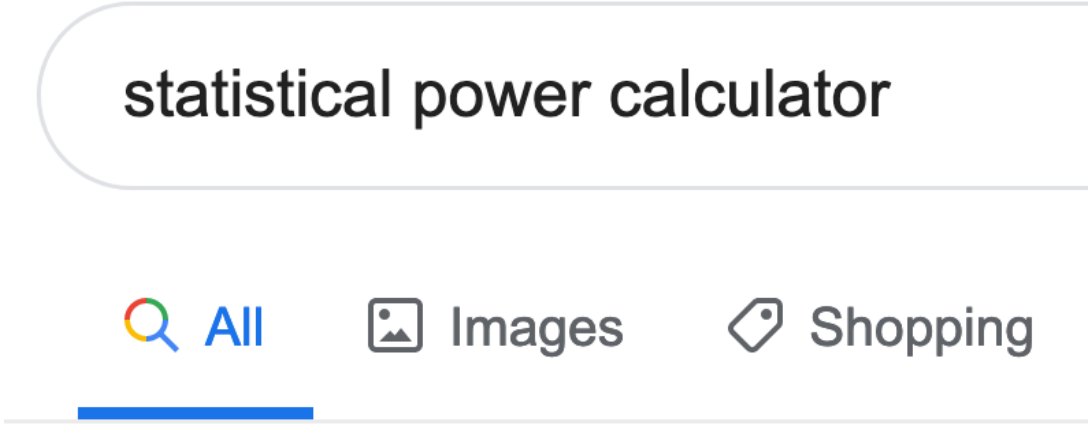


Enroll 200 participants



What's the right sample size?

Use a statistical power calculator to make sure you can potentially detect an effect



Or you can use simulations!

A quick note on randomization inference

- Previous power calculations work well on **larger samples**.
- On smaller samples, usually we can't rely on the **central limit theorem**.

Randomization Inference

What are the chances of something happening given all possible randomization scenarios?

The Lady Tasting Tea

- There's a lady in England that claims she can always know whether **tea or milk was poured first** in a cup.
- So you decide to **test it**.

- This is only one of the random allocations you could have used. There are other **19 combinations**!
  - Number of possible randomization allocations (or permutations): $\binom{n}{k}$

The Lady Tasting Tea

Possible Treatment Allocations

cup 1	cup 2	cup 3	cup 4	cup 5	cup 6
1	1	1	0	0	0
1	0	1	1	0	0
1	0	0	1	1	0
1	0	0	0	1	1
1	1	0	1	0	0
1	1	0	0	1	0
1	1	0	0	0	1
1	0	1	0	1	0

The Lady Tasting Tea

- If the woman is telling the truth, then she needs to pick out the 3 cups of tea with milk first.

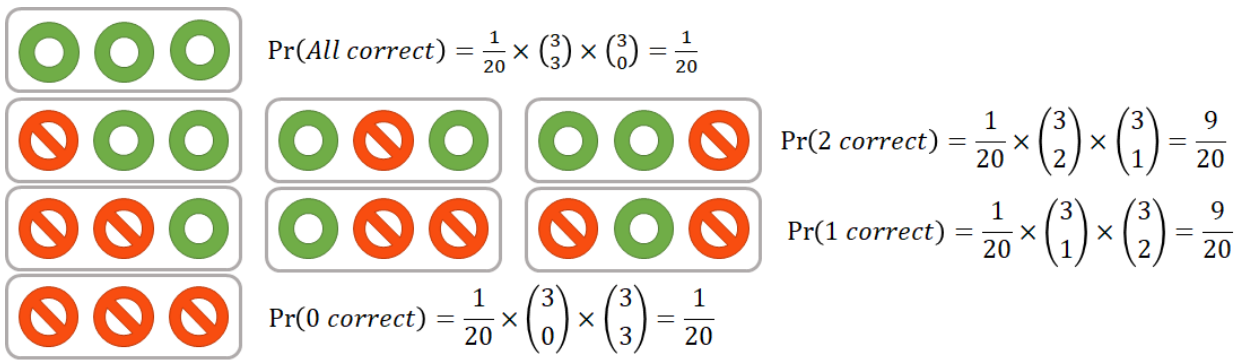
She does it!

What are the chances?

The Lady Tasting Tea

- If she was just guessing, she is equally likely to choose any of the **20 possibilities** (because $\binom{6}{3} = 20$).
- **Under the null hypothesis**, the probability of her choosing k correct ones out of n cups is:

$$Pr(Successes = k) = \frac{1}{Total\ Perm}\binom{n}{k}\binom{n}{n-k}$$



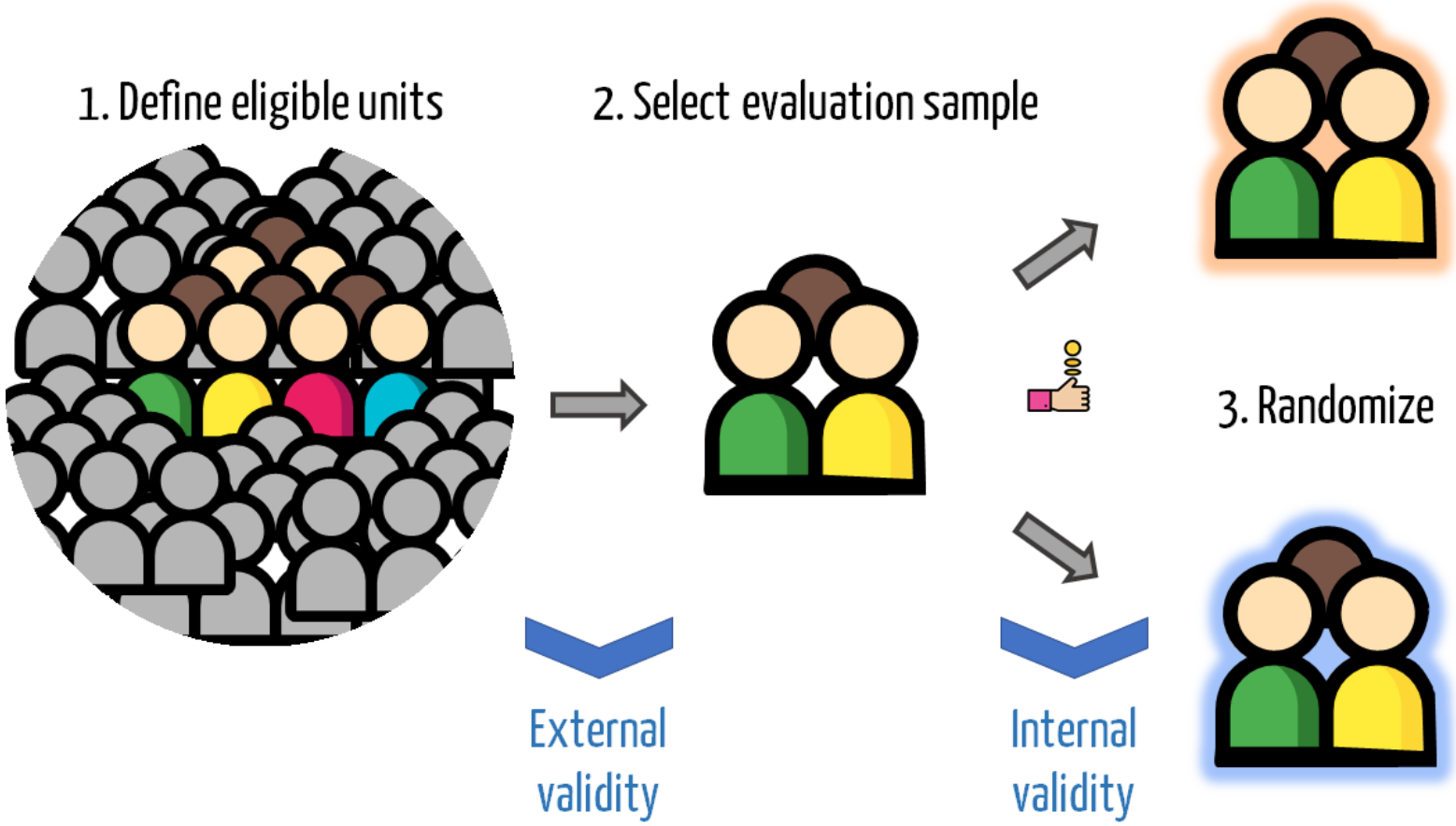
p-value = 0.05

Limitations and Potential Problems in Randomized Controlled Trials

Generalizibility of RCTs

- External Validity vs Internal Validity:
  - **External validity**: "The extent to which results can be generalized to other contexts or populations."
  - **Internal validity**: "[T]he extent to which the observed results represent the truth in the population we are studying."

External vs Internal Validity



- Many times, RCTs use **convenience samples**

Noncompliance and Attrition

Attrition

Units fall out of your sample

- **Can you give an example?**

Noncompliance and Attrition

Attrition

Units fall out of your sample

Noncompliance

Units that where assigned to one group end up in another

- **Example?**

Noncompliance and Attrition

- If **attrition** is correlated with the treatment, we're in trouble.

WHY?

Noncompliance and Attrition

- If **attrition** is correlated with the treatment, we're in trouble.

The ignorability assumption can break!

Noncompliance and Attrition

- If **attrition** is correlated with the treatment, we're in trouble.

The ignorability assumption can break!

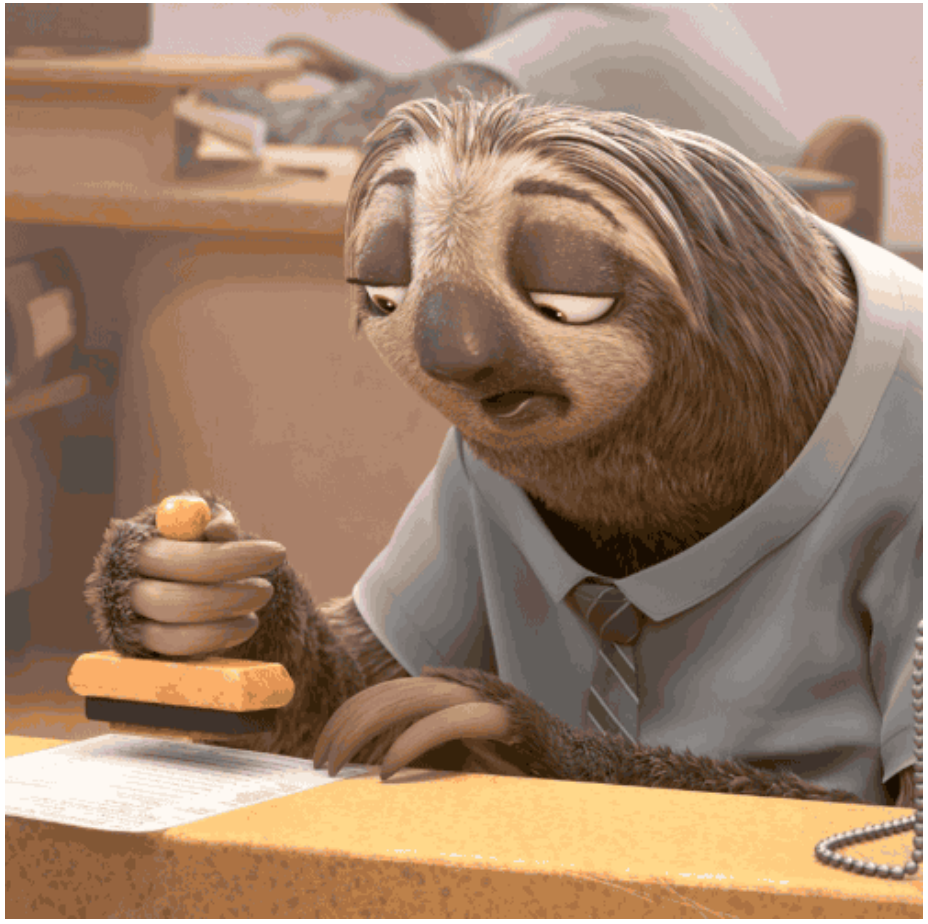
- **Noncompliance** is a problem for identifying ATE.
  - E.g. Treatment *assingment* is random, but not necessarily the *take-up* of the treatment
  - Stuck with **Intention to Treat (ITT) effect**.

Other problems to look out for

1. **Hawthorne effects**:
    - Effect that occurs when people behave differently because they are being watched
2. **John Henry effects**:
    - Effect that occurs when the control group works harder because they were not assigned to treatment.

Feasibility

- RCTs can be **expensive** and also **slow**.
- Organizations might **not be willing to randomize**.



Getting around some issues

- **Staggered adoption**: Eventually treat everyone, but at different times.
  - You can think about this when you have panel data.
- **Randomize in the bubble**:
  - Not randomize everyone, but maybe those that are on the margin of receiving the treatment.

Staggered adoption

	Time Period		
	1	2	3
Group 1			
Group 2			
Group 3			
Group 4			

	Time Period		
	1	2	3
Group 1			
Group 2			
Group 3			
Group 4			

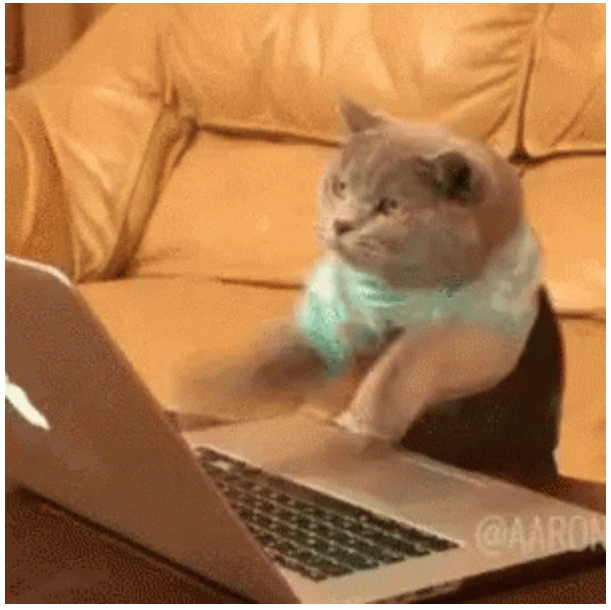
Staggered adoption

	Time Period		
	1	2	3
Group 1			
Group 2			
Group 3			
Group 4			

	Time Period		
	1	2	3
Group 1			
Group 2			
Group 3			
Group 4			

A/B Testing

A/B Testing



- A/B testing is very popular right now to test the effect of **small changes** in products on an outcome of interest.
- E.g. In web development, user would be exposed to **different versions of the same website** (only with subtle changes).
- Companies can quickly analyze the data, adapt their design, and continue testing features.

A/B Testing Jargon

- There's a lot of jargon in A/B testing that is thrown out there...

What does it all mean?

Conversions

Regret

Multi-armed bandits/ Contextual bandits

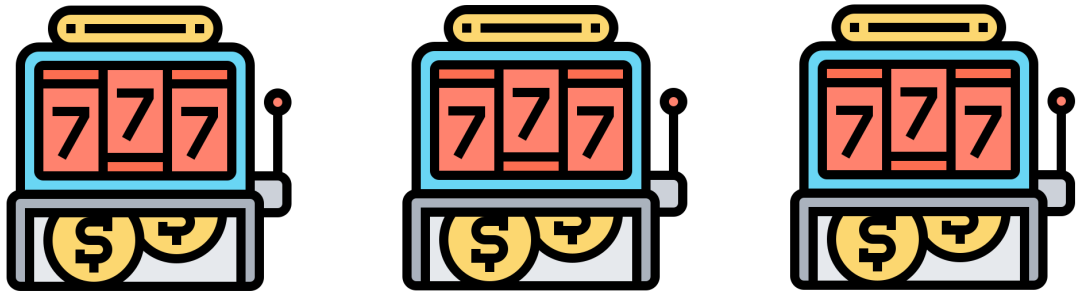
A/B Testing Jargon

- **Conversions**: "Cause the customer to take action"
- **Regret**: Loss of conversion due to a low-performing treatment.

Maximize conversions and minimize regret

A/B Testing Jargon

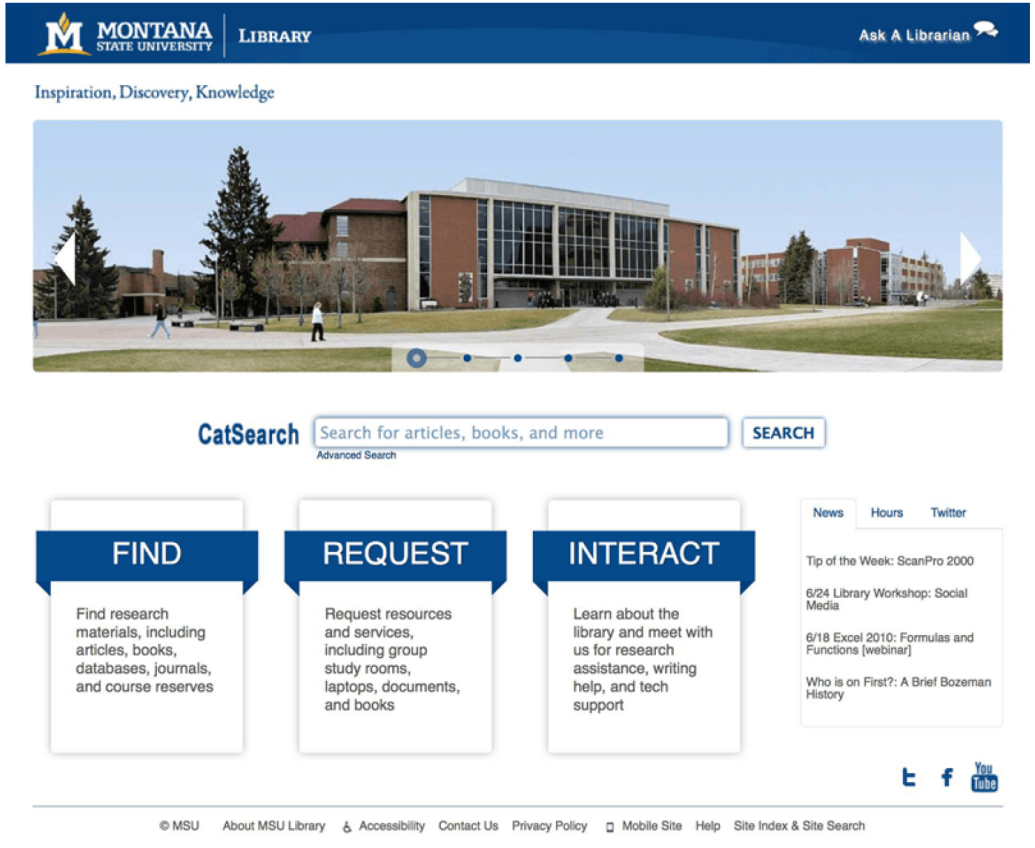
- **Multi-armed bandits/ Contextual bandits**
  - Name comes from machines at casinos.
  - You want to maximize total payout → Balance between exploration and exploitation



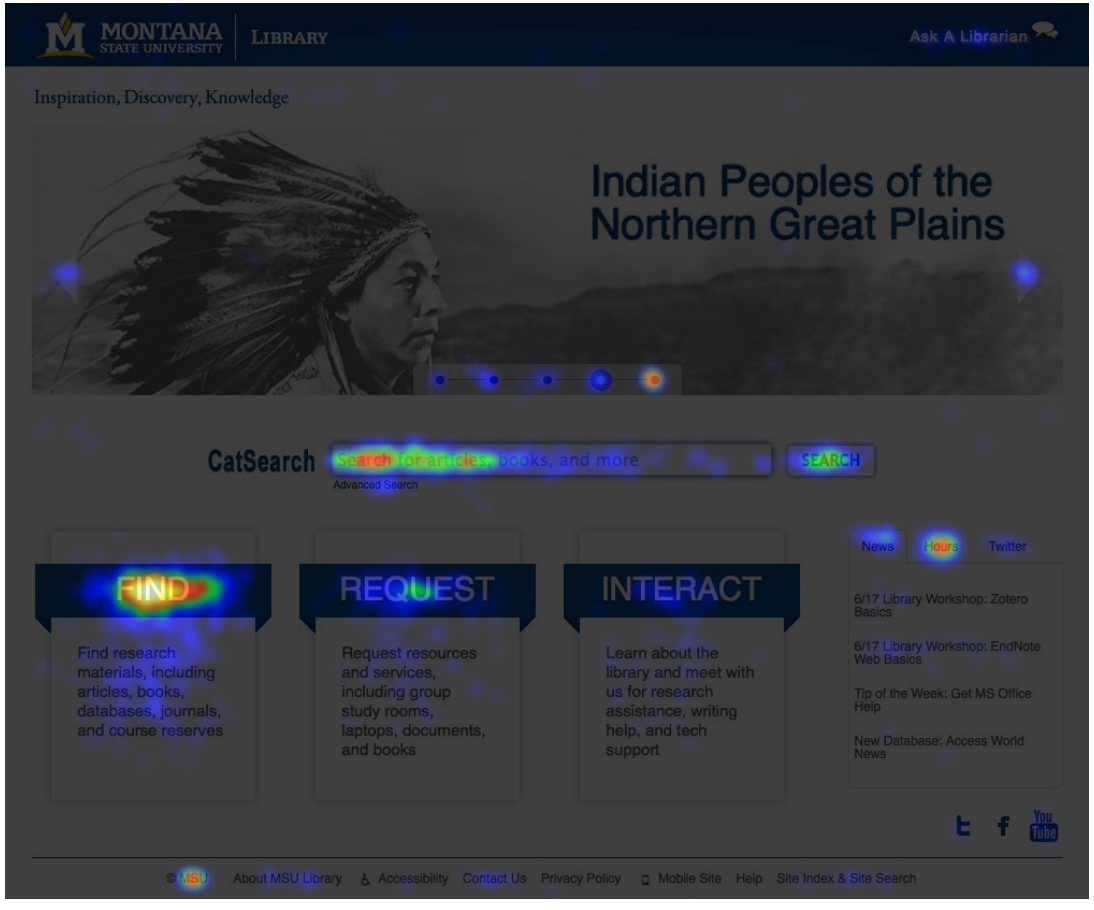
- **Multi-armed bandits** redirect users to the more successful versions of the treatment.
- A **contextual bandit** uses prior information from the user to select an action.

A/B Testing in Data: MSU Library

- Montana University Library website: Low interaction with the "Interact" Category

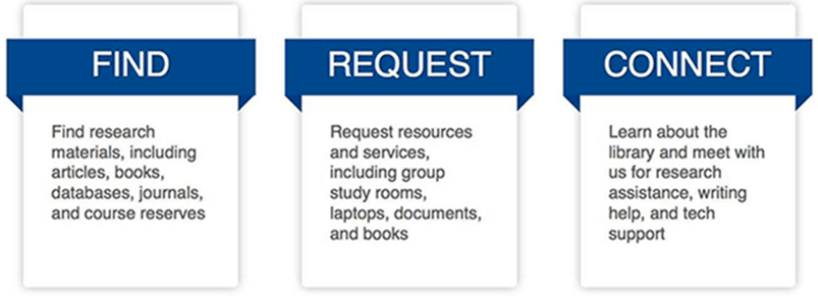


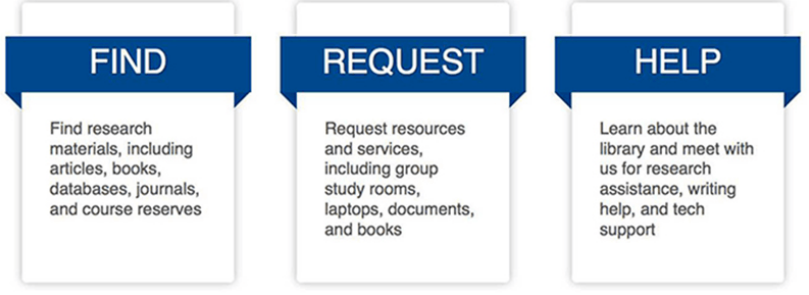
A/B Testing in Data: MSU Library

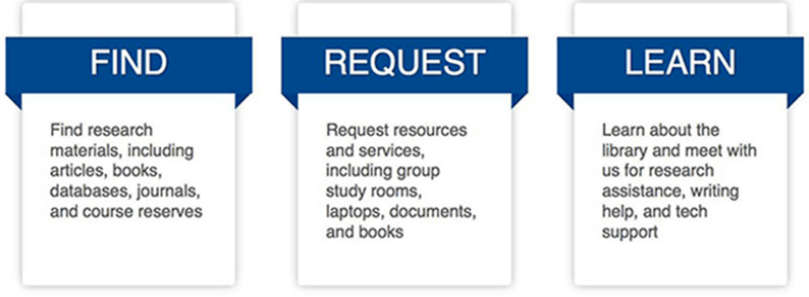


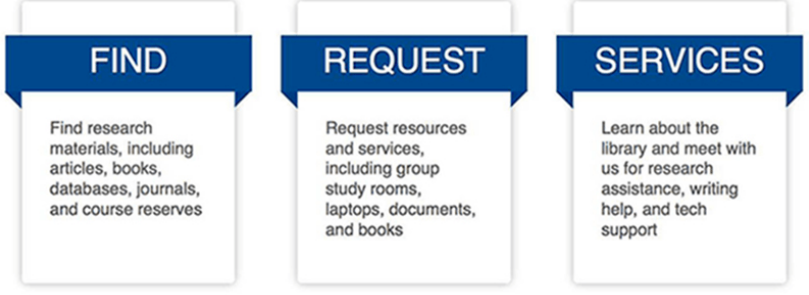
A/B Testing in Data: Different treatments

- Test four different names for that category









A/B Testing in Data: Analyze results

How would you analyze these results?

Homepage - Interact III (CrazyEgg)

May 29, 2013 - Jun 18, 2013

Stopped | Filters applied | View settings

Conversions / Page Metrics

All Sessions
37.45%

+ Add Segment

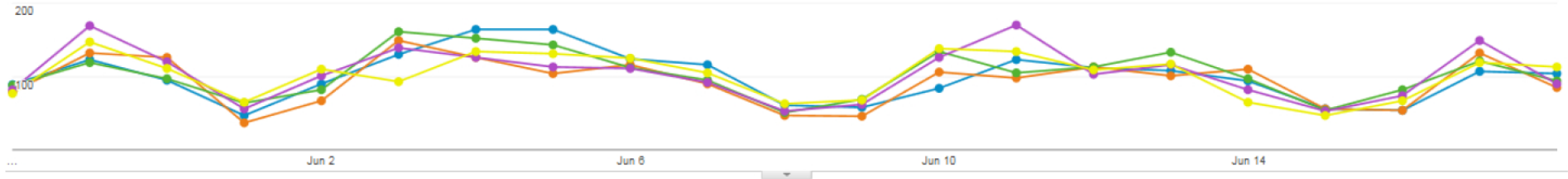
Explorer

Pageviews vs. Select a metric

Experiment stopped before winner was found

Day Week Month

Original - Interact Variation 1 - Connect Variation 2 - Learn Variation 3 - Help Variation 4 - Services



10,568 Pageviews

21 days of data

100% users included

Status:

Not enough data to identify winner

Primary Dimension: Variation

Plot Rows

Variation	Pageviews	Unique Pageviews	Avg. Time on Page	Entrances	Bounce Rate	% Exit	Page Value
Original - Interact	2,103	1,638	00:00:56	1,500	40.13%	36.38%	$0.00
Variation 3 - Help	2,187	1,690	00:01:39	1,633	50.03%	44.86%	$0.00
Variation 2 - Learn	2,166	1,693	00:01:14	1,612	46.90%	42.80%	$0.00
Variation 4 - Services	2,138	1,697	00:01:10	1,639	48.44%	44.15%	$0.00
Variation 1 - Connect	1,974	1,514	00:01:06	1,465	43.14%	39.21%	$0.00

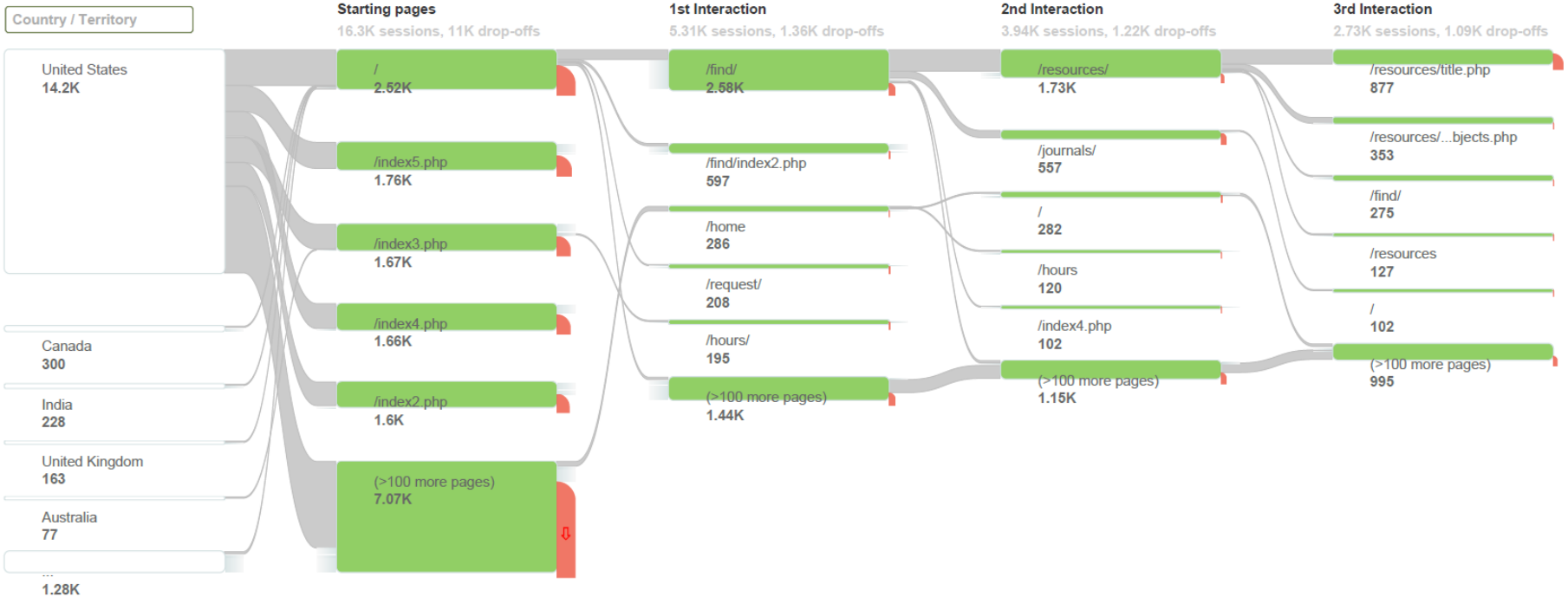
Show rows: 10 Go to: 1 1 - 5 of 5

Users Flow

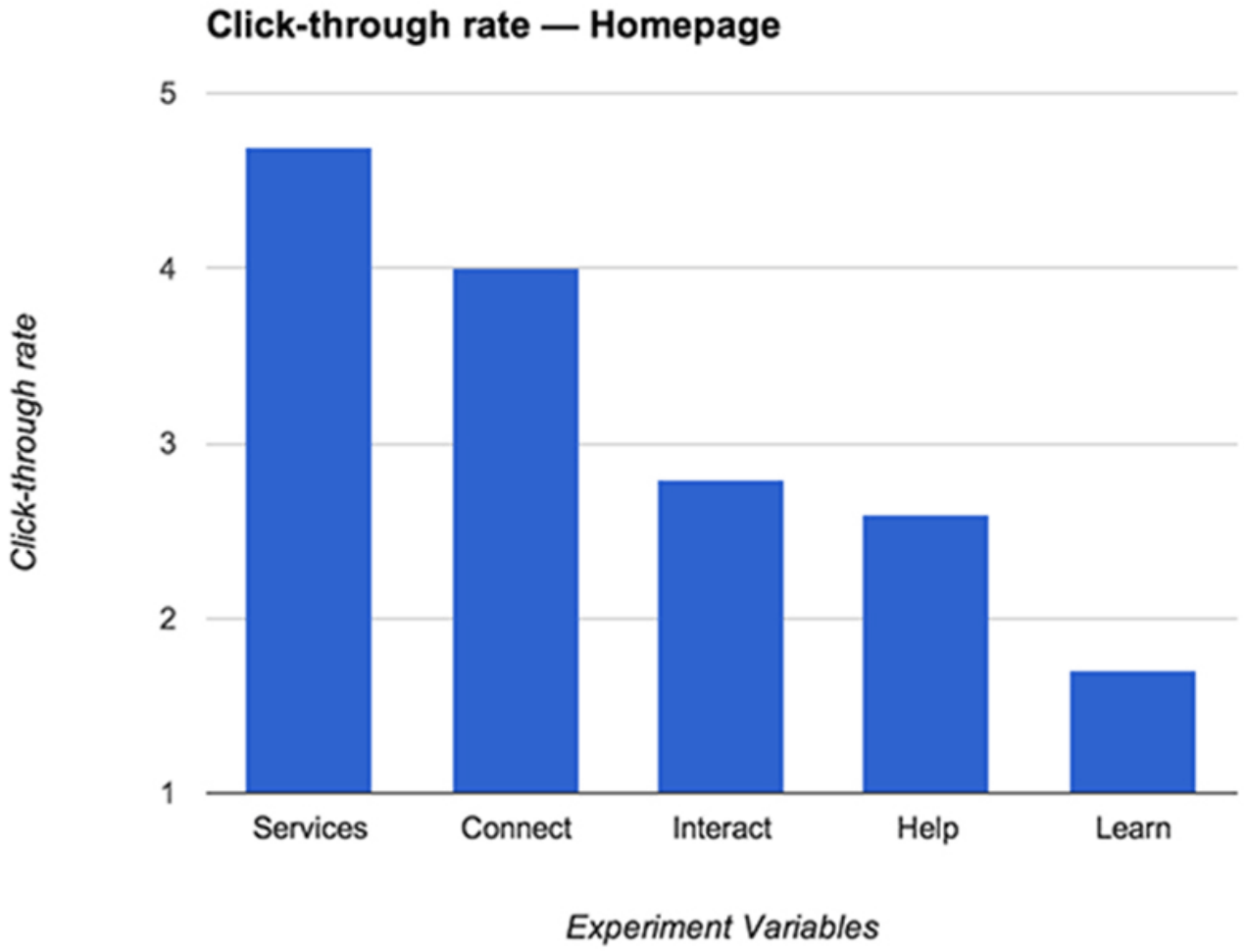
May 29, 2013 - Jun 18, 2013

All Sessions
100.00%

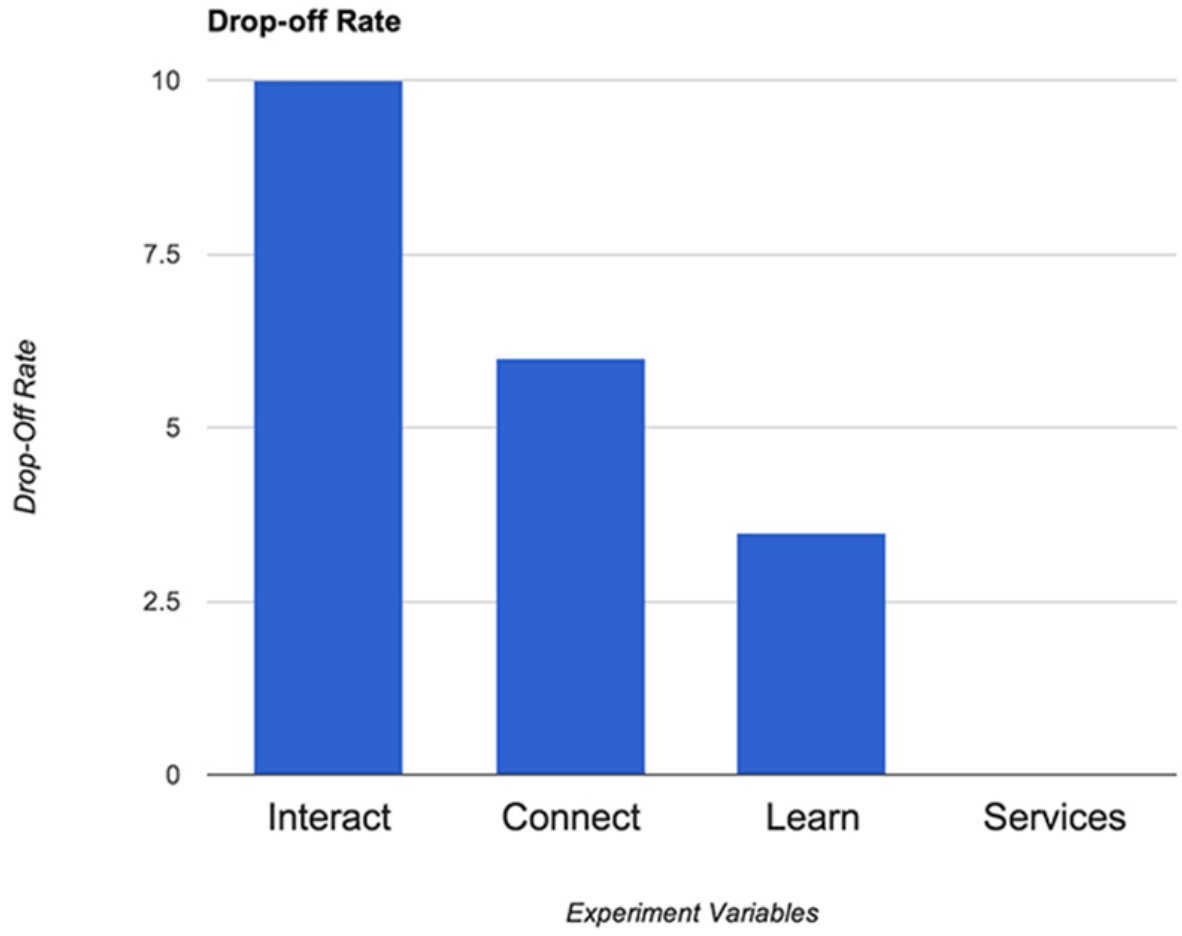
+ Add Segment



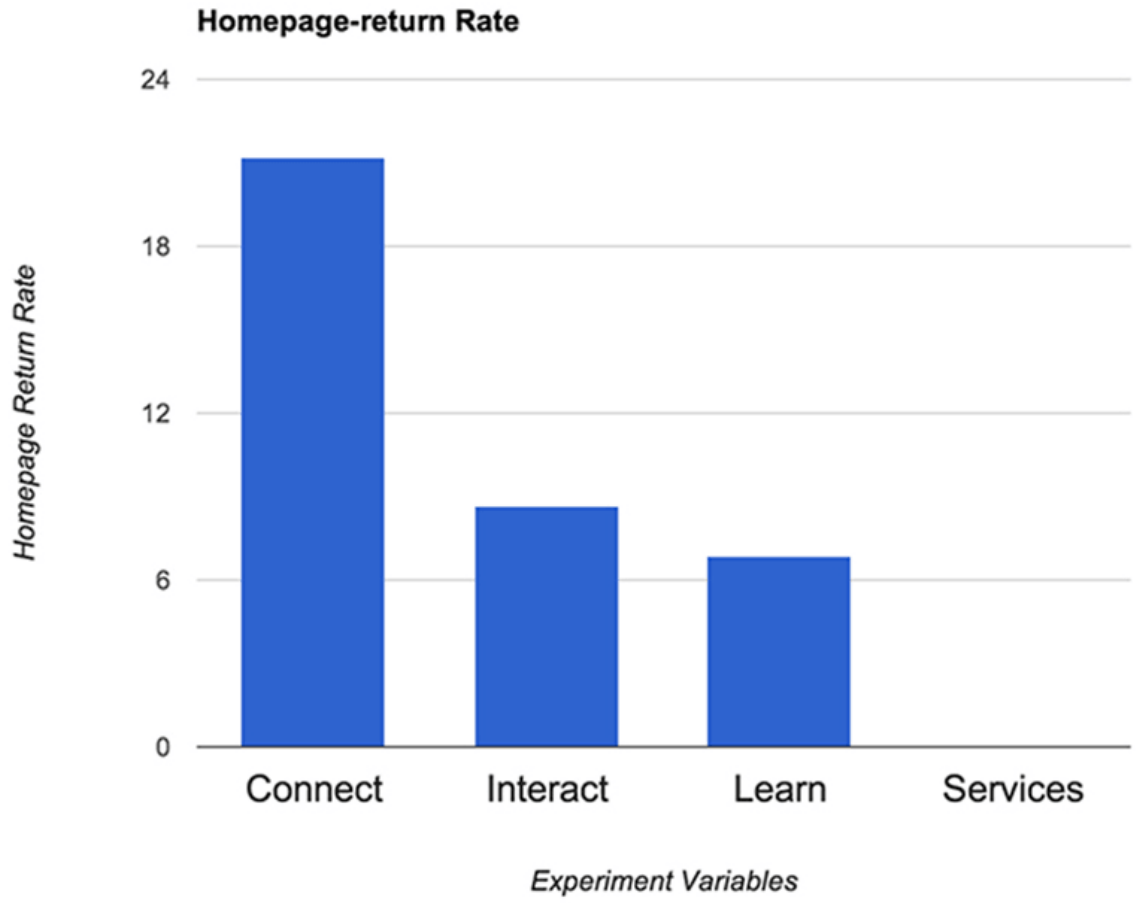
Homepage click-through



Drop-off


Drop-off Rate
10
7.5
5
2.5
0
Drop-Off Rate
Interact
Connect
Learn
Services
Experiment Variables

Homepage return



Wrapping things up

- Randomized controlled trials are great... **but not for everything!**
- Randomization buys us **no systematic selection on observables or unobservables**
    - But things can go wrong, too!

Check your assumptions and look out for potential issues!

Next class

- Introduction to **observational studies**
- Selection on **observables**
- The wonderful world of **matching!**

References

- Angrist, J. and S. Pischke. (2015). "Mastering Metrics". *Chapter 1*.
- Heiss, A. (2020). "Program Evaluation for Public Policy". *Class 7: Randomization and Matching, Course at BYU*
- Imbens, G. and D. Rubin. (2015). "Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction". *Chapter 1*
- Young, S. (2014). "Improving Library User Experience with A/B Testing: Principles and Process". *Weave. Vol 1, Issue 1*.