Bush 631-603: Quantitative Methods

Lecture 2 (01.25.2022): Causality vol. I

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

Spring 2022

What is today's plan?

- Causality and deriving cause-effect relationship.
- Research designs to assess causality.
- Randomized controlled experiments (RCTs).
- R work: more ways to learn of our data, sub-setting data, factor variables.

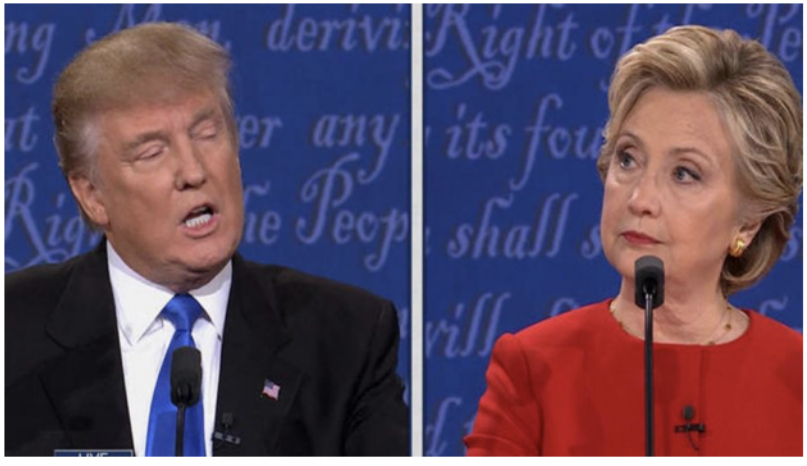
Causality

- Identify causes for outcomes of interest:
  1. Universal health care and better health status among poor.
  2. Drop in president approval during war.
- Establish causality:

$$\text{Cause} \rightarrow \text{Effect}$$

Establish causality

2016 turnout: 59.2% of VEP
2020 turnout: 62% of VEP



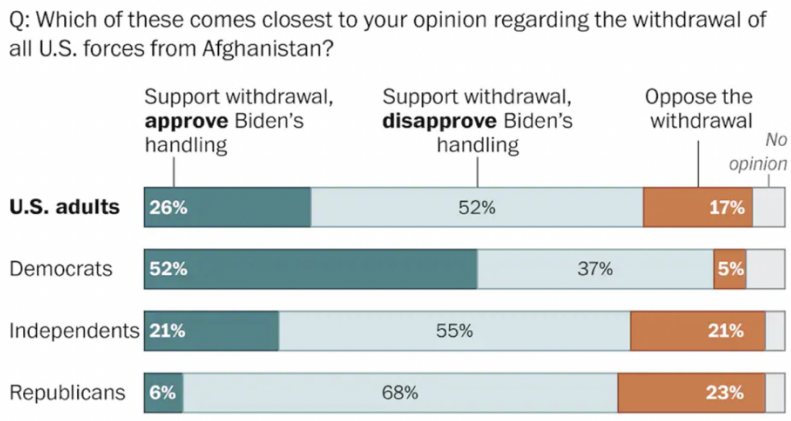
- Candidate gender $\rightarrow$ election turnout ?

Experiments

- Test causal effects using hypothetical scenario.
- Some use actual setting (natural experiment).
- Candidate gender and public support? use an experiment...

Experiments

President party $\rightarrow$ foreign policy?



Experiments, how?

- Test causal effects using a *treatment*.
- Treatment(s) represents the proposed causal factor(s).
- Manipulate treatments - assign different values.
- Measure and compare outcome across treatments.

Experiments in FP

Hawks – Doves and Foreign Policy Reconciliation

Mattes and Weeks (2019)

The design

- Elements of experiment:
  - Hypothetical scenario.
  - Adversary: China.
  - Important FP issue - access to arctic.
  - Outcome measured: approval of president's actions.
- Treatments:
  - Description of factors.
  - Vary between groups.

How does it look like?

-Background information:

"The year is 2027. The U.S. President is John Richards. President Richards took office in 2025 after serving in the U.S. Senate for six years."

How does it look like?

- The leader's type (variable name = hawk_t):

Hawk/Dove	
Hawk	**Dove**
. . . has a reputation for favoring military solutions over diplomatic ones. He has repeatedly emphasized that military force is essential to protecting American national security. President Richards says that he will not shy away from using force where necessary. He has long said that "the only way to achieve peace is to be ready for war."	. . . has a reputation for favoring diplomatic solutions over military ones. He has repeatedly emphasized that military force is not the answer to protecting American national security. President Richards says that he believes in diplomacy and negotiations and will use military force only as a last resort. He has long said that "the only way to achieve peace is to act peacefully."

How does it look like?

- The setting (all respondents):
  - China: distrusted adversary.
  - Tense relations.
  - Specific issue - access to arctic.

"One very tense issue is access to the Arctic. The Arctic contains up to 40 percent of the world's oil and gas resources and provides vital shipping routes between continents. In 2027, the U.S. and China both have a major military presence in the Arctic. Each country has thousands of troops in the area and holds frequent military exercises in the region."

How does it look like?

- President Richards and China:

"In his 2027 State of the Union speech, President Richards declares that getting China to cooperate is important for achieving U.S. foreign policy goals."

- Policy choice (variable name = rapproche_t)

Policy Choice	
Conciliatory	**Status Quo**
. . . announces that he is sharply reducing the U.S. military presence in the Arctic. He is withdrawing a third of the U.S. forces currently in the Arctic and is calling off planned military exercises in the region.	. . . announces that he is maintaining the current U.S. military presence in the Arctic. He will continue to keep U.S. forces in the Arctic and will carry through with planned military exercises in the region.

How does it look like?

- Measuring outcomes:
  1. President approval (variable name = hddv1): rate on a 1-5 scale.
  2. Trust: level of international trust in other nations (yes/no).
  3. Internationalism: US involvement in world affairs (1-4 scale).
- Respondents' characteristics:
  1. Gender.
  2. Voted in 2016?

The experiment data

```
dim(mydata)
```

```
## [1] 1199   32
```

```
head(mydata)
```

```
## # A tibble: 6 x 32
##       caseid  hawk_t party_t rapproche_t success_t    hawk    intl   trust vo
##        <dbl> <dbl+l> <dbl+l>   <dbl+lbl> <dbl+lbl> <dbl+l> <dbl+l> <dbl+l> <d
## 1 329144398 2 [Dip~ 1 [Rep~ 1 [Reducin~ 1 [Pulls~ 4 [Agr~ 4 [Agr~ 2 [The~ 4
## 2 329105048 1 [Mil~ 1 [Rep~ 1 [Reducin~ 1 [Pulls~ 2 [Dis~ 4 [Agr~ 2 [The~ 4
## 3 328964530 1 [Mil~ 2 [Dem~ 1 [Reducin~ 1 [Pulls~ 2 [Dis~ 2 [Dis~ 2 [The~ 4
## 4 329130310 1 [Mil~ 2 [Dem~ 2 [Maintai~ 2 [Addit~ 4 [Agr~ 5 [Agr~ 1 [The~ 4
## 5 328809639 1 [Mil~ 2 [Dem~ 2 [Maintai~ 1 [Pulls~ 3 [Nei~ 2 [Dis~ 2 [The~ 1
## 6 329124511 2 [Dip~ 1 [Rep~ 2 [Maintai~ 2 [Addit~ 3 [Nei~ 4 [Agr~ 2 [The~ 4
## # ... with 23 more variables: polact_1 <dbl+lbl>, polact_2 <dbl+lbl>,
## #   polact_3 <dbl+lbl>, polact_4 <dbl+lbl>, hddv1 <dbl+lbl>,
## #   hdmed1_strat <dbl+lbl>, hdmed1_pacifist <dbl+lbl>,
## #   hdmed1_warmonger <dbl+lbl>, hddv2 <dbl+lbl>, hdmed2_strat <dbl+lbl>,
## #   hdmed2_pacifist <dbl+lbl>, hdmed2_warmonger <dbl+lbl>, birthyr <dbl>,
## #   gender <dbl+lbl>, educ <dbl+lbl>, pid3 <dbl+lbl>, pid7 <dbl+lbl>,
## #   ideo5 <dbl+lbl>, newsint <dbl+lbl>, pew_religimp <dbl+lbl>, ...
```

The experiment data

summary(mydata)

Console | Terminal | R Markdown | Markers | Jobs

~/

```
     hddv2        hdmed2_strat   hdmed2_pacifist hdmed2_warmonger    birthyr         gender           educ            pid3
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000    Min.   :1925   Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:1.000    1st Qu.:1956   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000
 Median :4.000   Median :4.000   Median :3.000   Median :2.000    Median :1971   Median :2.000   Median :3.000   Median :2.000
 Mean   :3.505   Mean   :3.403   Mean   :2.716   Mean   :2.304    Mean   :1970   Mean   :1.549   Mean   :3.355   Mean   :2.108
 3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:3.000    3rd Qu.:1984   3rd Qu.:2.000   3rd Qu.:5.000   3rd Qu.:3.000
 Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000    Max.   :1998   Max.   :2.000   Max.   :6.000   Max.   :5.000

      pid7            ideo5         newsint       pew_religimp    approve_b                      internatiolism   ally_trust
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   :0.0000   Agree Somewhat            :450   Min.   :0.0000
 1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.00   1st Qu.:0.0000   Agree Strongly            :112   1st Qu.:1.0000
 Median :4.000   Median :3.000   Median :1.000   Median :2.00   Median :1.0000   Disagree Somewhat         :288   Median :1.0000
 Mean   :3.642   Mean   :3.069   Mean   :1.753   Mean   :2.32   Mean   :0.6415   Disagree Strongly         :105   Mean   :0.7922
 3rd Qu.:6.000   3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.:4.00   3rd Qu.:1.0000   Neither Agree nor Disagree:244   3rd Qu.:1.0000
 Max.   :8.000   Max.   :6.000   Max.   :7.000   Max.   :4.00   Max.   :1.0000                                    Max.   :1.0000
                                                                NA's   :231                                       NA's   :1
```

Exploring the data: cross-tabs

```
# Cross-tabs
table(type = mydata$hawk_t, support = mydata$hddv1)
```

```
##      support
## type   1   2   3   4   5
##    1  59 132 148 187  74
##    2  73  83  83 217 143
```

```
tab2 <- table(support = mydata$hddv1, party = mydata$party_t)
addmargins(tab2)
```

```
##         party
## support    1    2  Sum
##     1     55   77  132
##     2    100  115  215
##     3    115  116  231
##     4    209  195  404
##     5    120   97  217
##     Sum  599  600 1199
```

Data analysis: first steps

```r
# Calculate mean support for president
# Using the $ sign method
mean1 <- sum(mydata$hddv1) / nrow(mydata)
mean1
```

```
## [1] 3.299416
```

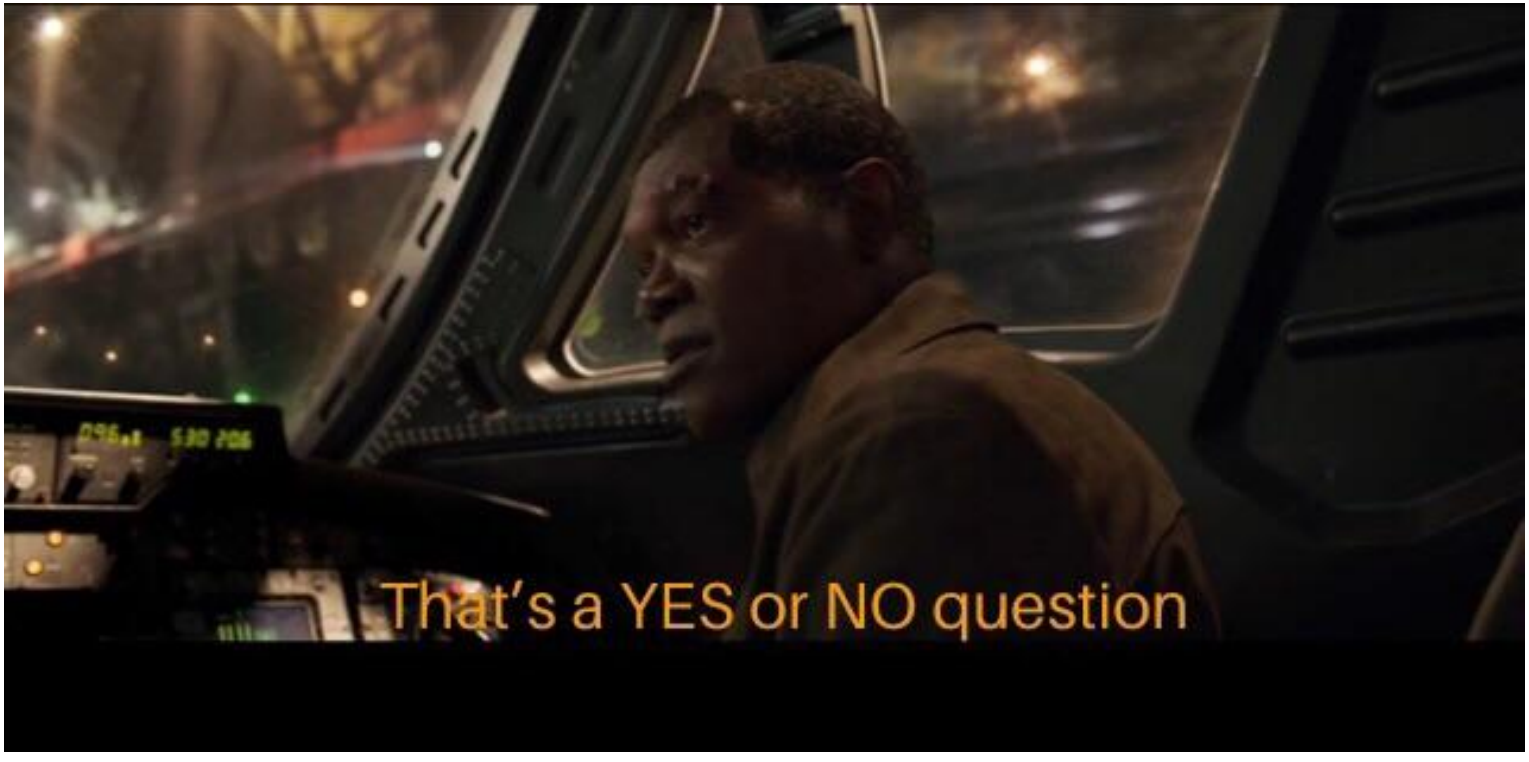
```r
# Using the indexing method
mean2 <- sum(mydata[,14]) / nrow(mydata)
mean2
```

```
## [1] 3.299416
```

```r
# Mean proportion of support
mean3 <- mean(mydata$approve_b, na.rm = TRUE)
mean3
```

```
## [1] 0.6415289
```

Logical values



- TRUE / FALSE output.

Logical values

```
class(FALSE)
```

```
## [1] "logical"
```

```
as.integer(TRUE)
```

```
## [1] 1
```

```
v1 <- c(FALSE,TRUE,TRUE,FALSE,FALSE)

mean(v1)
```

```
## [1] 0.4
```

```
sum(v1)
```

```
## [1] 2
```

Relational operators

- Evaluate the relationship between two values.
- Results are displayed as logical values

```
12 > 9
```

```
## [1] TRUE
```

```
"aggies" == "Aggies"
```

```
## [1] FALSE
```

```
"Aggies" == "Aggies"
```

```
## [1] TRUE
```

Relational operators

- Apply to vectors: results are logical values.

```
v3 <- c(4,8,-1,-9,7)

v3 < 0
```

```
## [1] FALSE FALSE  TRUE  TRUE FALSE
```

```
v3 >= 4
```

```
## [1]  TRUE  TRUE FALSE FALSE  TRUE
```

```
v3 != 7
```

```
## [1]  TRUE  TRUE  TRUE  TRUE FALSE
```

Sub-setting data

- Partition/split our data for certain calculations.

```
# Proportions of support by party
mean(mydata$approve_b[mydata$party_t == 1], na.rm = TRUE)
```

```
## [1] 0.6797521
```

```
mean(mydata$approve_b[mydata$party_t == 2], na.rm = TRUE)
```

```
## [1] 0.6033058
```

```
# Mean approval score by party
mean(mydata$hddv1[mydata$party_t == 1], na.rm = TRUE)
```

```
## [1] 3.398998
```

```
mean(mydata$hddv1[mydata$party_t == 2], na.rm = TRUE)
```

```
## [1] 3.2
```

Sub-setting data

- Create subset of one group only.
- Only 'Hawkish' presidents.

```
# Sub-set 'hawks'
mysubdata1 <- mydata[mydata$hawk_t == 1,]
dim(mysubdata1)
```

```
## [1] 600  32
```

```
# Calculate mean support/approval
mean(mysubdata1$hddv1)
```

```
## [1] 3.141667
```

```
mean(mysubdata1$approve_b, na.rm = TRUE)
```

```
## [1] 0.5774336
```

Sub-setting data

- Subset function: construct a dataset only for the variables we are interested in.

```
mysubdata2 <- subset(mydata, subset = (hawk_t == 2))
```

View(mysubdata2)

	caseid Case ID	hawk_t hawk_t	party_t party_t	rapproche_t rapproche_t	success_t success_t	hawk hawk	intl internationalism	trust trust	voted16 Voted in 2016
1	329144398	2	1	1	1	4	4	2	4
2	329124511	2	1	2	2	3	4	2	4
3	329023155	2	2	1	1	2	4	1	4
4	329124618	2	2	1	2	5	1	2	4
5	329011534	2	1	2	2	1	3	2	4
6	329056352	2	1	1	1	1	2	2	4
7	328905656	2	2	1	1	2	4	1	1
8	329147372	2	2	2	2	5	1	1	4
9	329147427	2	1	2	1	2	4	2	4
10	329147506	2	1	1	2	4	1	2	4
11	328849981	2	1	2	2	5	4	2	4
12	329002390	2	2	2	2	4	1	2	4
13	328770388	2	2	1	1	3	4	2	4
14	329231291	2	1	2	2	4	3	2	1
15	329254548	2	1	2	2	3	4	2	4

Calculating Group means

- Evaluate difference in support between Republican - Democrat president

```
# Create sub-samples for rep/dem president
mysubdata_rep <- subset(mydata, subset = (party_t == 1))
mysubdata_dem <- subset(mydata, subset = (party_t == 2))
```

```
# Compute difference in means
mean(mysubdata_rep$approve_b, na.rm = TRUE) -
  mean(mysubdata_dem$approve_b, na.rm = TRUE)
```

```
## [1] 0.07644628
```

Compare means within a subset

- Hawkish president: sub-sets for respondents' gender

```
# Create sub-samples for gender president
mysubdata4_male <- subset(mydata, subset = (hawk_t == 1 & gender == 1))
mysubdata4_female <- subset(mydata, subset = (hawk_t == 1 & gender == 2

# Compute difference in means
mean(mysubdata4_female$approve_b, na.rm = TRUE) -
  mean(mysubdata4_male$approve_b, na.rm = TRUE)
```

```
## [1] -0.06519359
```

Conditional Statements

- Useful for **creating new variables (columns)** in our data.
- Conditional - based on values of other variables.
- Rely on TRUE / FALSE logical statements.

The ifelse() function

- ifelse(condition, value if TRUE, value if FALSE).
- Can accept multiple conditions.

```
# Create variable based on conditions
mydata$new1 <- ifelse(mydata$voted16 == 4 &
                          mydata$gender == 2,1,0)

# Cross-tabs: variable values
table(women_voters = mydata$new1)
```

```
## women_voters
##   0   1
## 683 516
```

```
# Cross-tabs: counts of support for new variable
table(newvar = mydata$new1, support = mydata$approve_b)
```

```
##       support
## newvar   0   1
##      0 186 361
##      1 161 260
```

The ifelse() function

- Respondents' level of 'hawkishness' (survey item):

> *hawk* – hawkishness, measured based on agreement with the statement "The use of military force only makes problems worse." 1 = Disagree strongly, 2 = Disagree somewhat, 3 = Neither agree nor disagree, 4 = Agree somewhat, 5 = Agree strongly

```
# Create variable
mydata$no_hawks <- ifelse(mydata$hawk>3,1,0)

# Cross-tabs: variable values
table(NoHawks = mydata$no_hawks)
```

```
## NoHawks
##   0   1
## 757 442
```

Factor Variables

- Categorical variable with finite number of distinct levels/values.

intl – internationalism, measured based on agreement with the statement "The United States needs to play an active role in solving conflicts around the world." 1 = Disagree strongly, 2 = Disagree somewhat, 3 = Neither agree nor disagree, 4 = Agree somewhat, 5 = Agree strongly

Factor Variables

- Looking at factor variables

```
mydata$internatiolism <- as.factor(mydata$internatiolism)
class(mydata$internatiolism)
```

```
## [1] "factor"
```

```
levels(mydata$internatiolism)
```

```
## [1] "Agree Somewhat"              "Agree Strongly"
## [3] "Disagree Somewhat"           "Disagree Strongly"
## [5] "Neither Agree nor Disagree"
```

```
table(mydata$internatiolism)
```

```
##
##              Agree Somewhat                 Agree Strongly
##                         450                            112
##           Disagree Somewhat              Disagree Strongly
##                         288                            105
## Neither Agree nor Disagree
##                         244
```

tapply() command

- Apply a function across all levels of factor variable.
- Sort in desired order

```
# tapply: calculate mean approval for all levels
app_int <- tapply(mydata$approve_b, mydata$internatiolism, mean, na.rm = TRUE)
app_int
```

```
##              Agree Somewhat              Agree Strongly
##                   0.6544503                   0.6086957
##           Disagree Somewhat           Disagree Strongly
##                   0.6637555                   0.6547619
## Neither Agree nor Disagree
##                   0.5966851
```

```
# Sort by value
sort(app_int)
```

```
## Neither Agree nor Disagree              Agree Strongly
##                   0.5966851                   0.6086957
##              Agree Somewhat           Disagree Strongly
##                   0.6544503                   0.6547619
##           Disagree Somewhat
##                   0.6637555
```

Causal Inference

- The importance of **counterfactual**

Causal Inference

- *Counterfactual* in Foreign policy

May 2000: Israel withdraws from S. Lebanon
Prime minister: Ehud Barak

Causal Inference

- Gender / military experience as potential *causal factors*.
- Why maybe?
- Can we isolate the factors that lead to different outcomes?

Causal Inference: QSS textbook

- Does an applicant's race affects the chances she/he are offered a job?
- Race $\rightarrow$ causal factor for job prospects?
- My name affects my chances of landing a job.
- Is it only my name/race??

Causal Inference: QSS textbook

Résumé i	*Black-sounding name* T_i	*Callback* $Y_i(1)$	*Callback* $Y_i(0)$	*Age*	*Education*
1	1	1	?	20	college
2	0	?	0	55	high school
3	0	?	1	40	graduate school
⋮	⋮	⋮	⋮	⋮	⋮
n	1	0	?	62	college

Other factors:

- Age.
- Education.
- Other?

Causal Inference

The fundamental problem of causal inference

- We cannot observe counterfactual outcomes.
- Assume research design helps us infer about our observed counterfactual outcomes.
- Identification process: same situation, **one** factor changes (president's gender, applicant race).
- Not possible in reality: *immutable characteristics*.

Randomized Controlled Trails (RCTs)

The gold standard of causal inference

Why?

- Establish causality by *isolating* the factor of interest.

How?

- Randomization - random assignment to treatments.

RCTs

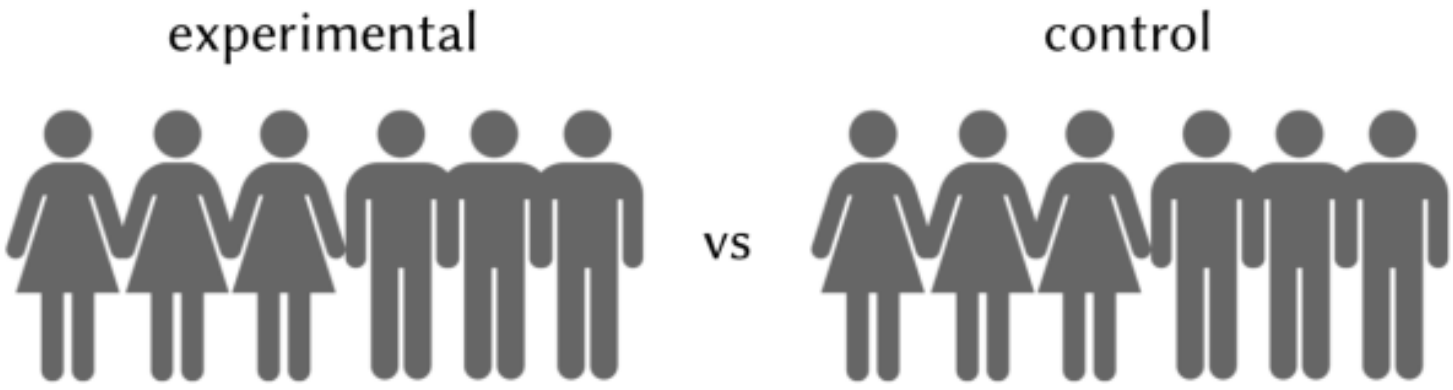
Sample Average Treatment Effect (SATE)

- The average individual-level treatment effect.
- Defined as:

$$SATE = 1/n \sum_{i=1}^{n} Y_i(1) - Y_i(0)$$

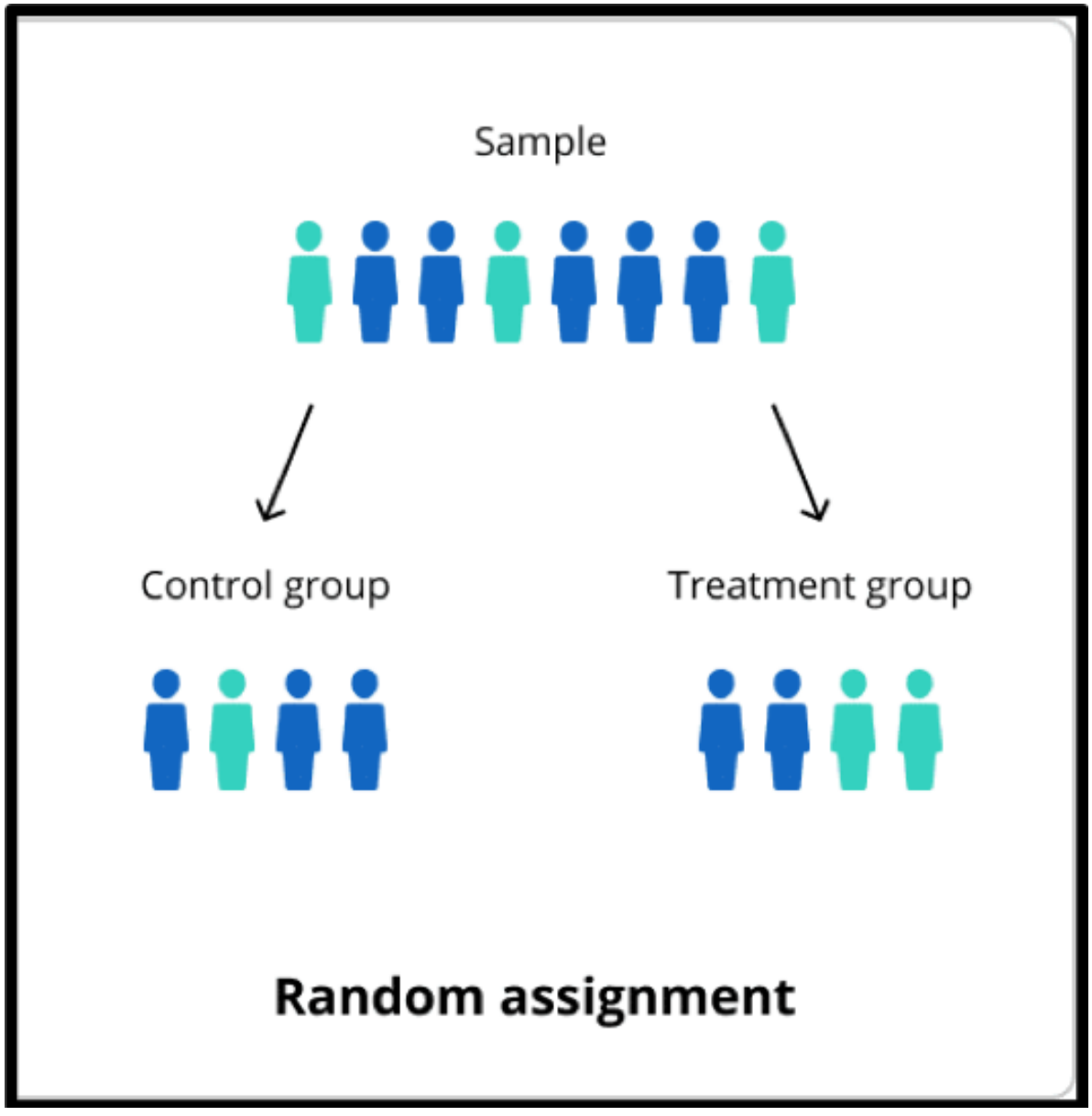
- n $\rightarrow$ sample size.
- i $\rightarrow$ respondent in the sample.

RCTs



- Compare outcomes between groups.
- SATE: average outcome between groups.
- *Difference-in-means estimator*

Random Assignment



Random Assignment

- Why important?
- Confounding factors similar in sample.
- Our treatment $\rightarrow$ Variation in outcomes.
- Our treatment $\rightarrow$ Causal factor.

Internal Validity

- Design satisfy causal assumptions?
- How well does the help us in testing our research question.
- Experiments offer strong *internal* validity.

External Validity

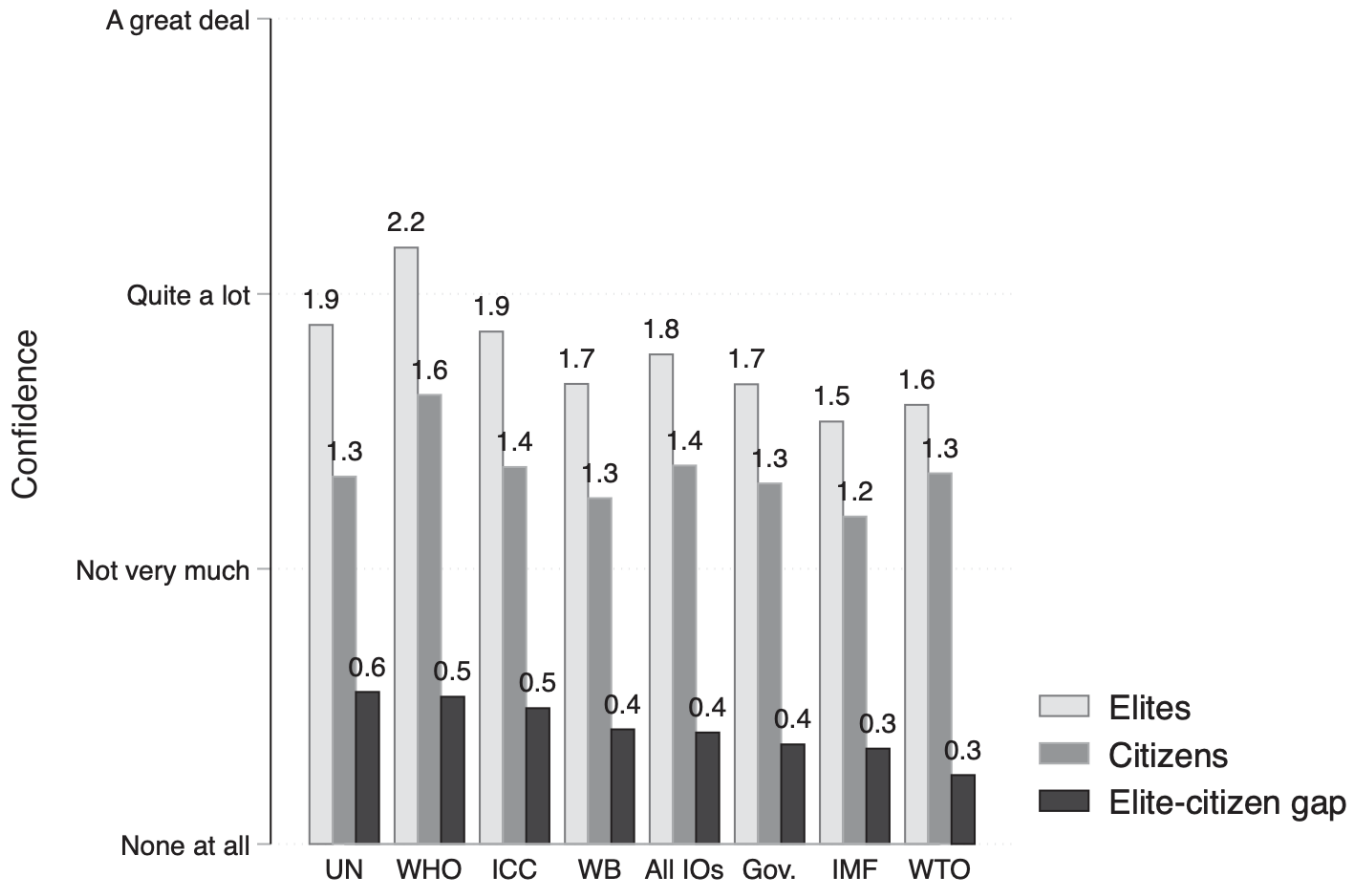
- Can we trust our results beyond the sample?
- Students sample = general public?
- *Convenient samples*: why?
- Main problem: *sample selection bias*
- Other design problems:
  1. Setting: lab versus real-world (Hawthorne effect).
  2. Unrealistic treatments: missing information.

Reduce external validity

- Replications - same design, vary the sample:
  - General public and special samples (students, elites, experts, etc.)
  - Cross-national.
  - Multiple samples of same population.
- Consistency in results $\rightarrow$ more confidence in proposed causal factor.

Replications

- Public and elite samples: legitimacy of IOs



Causal Inference and president approval

```
# Treatment = president type
# What is proportion of support
tapply(mydata$approve_b, mydata$hawk_t, mean, na.rm = TRUE)
```

```
##         1         2
## 0.5774336 0.6976744
```

```
# Treatment = president party
# What is proportion of support
tapply(mydata$approve_b, mydata$party_t, mean, na.rm = TRUE)
```

```
##         1         2
## 0.6797521 0.6033058
```

Causal Inference and president approval

- Grouping treatments by president party and policy choice

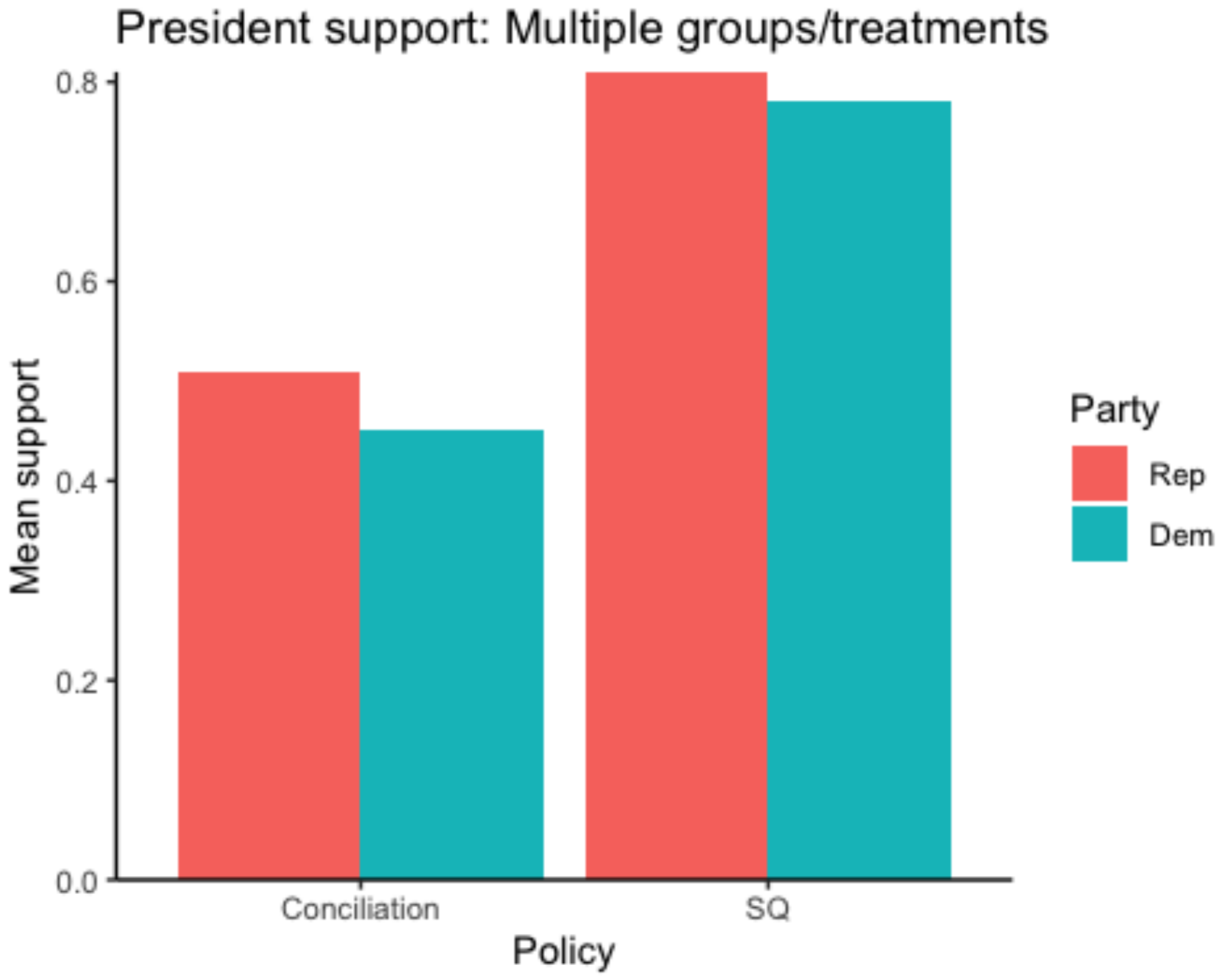
```
# Create factorial variable for policy and party
mydata$party_policy <- NA
mydata$party_policy[mydata$party_t == 1 & mydata$rapproche_t == 1] <- 1
mydata$party_policy[mydata$party_t == 1 & mydata$rapproche_t == 2] <- 2
mydata$party_policy[mydata$party_t == 2 & mydata$rapproche_t == 1] <- 3
mydata$party_policy[mydata$party_t == 2 & mydata$rapproche_t == 2] <- 4
```

```
# What is proportion of support by group
tapply(mydata$approve_b, mydata$party_policy, mean, na.rm = TRUE)
```

```
##         1         2         3         4
## 0.5138889 0.8134328 0.4492188 0.7763158
```

Causal Inference and president approval

- Grouping treatments by president party and policy choice
- Visual:



Wrapping up week 2

Causality vol. I:

- Assessing causal effects.
- Experimental designs (RCTs).
- Counterfactuals.
- Randomization.
- Internal and external validity.
- R work: cross-tabs, relational operations, sub-set data, ifelse(), factor variables.