Modulo 3

Experiments and Field Experiments

Umberto Mignozzetti

School of International Relations FGV
FGV

2020-07-06

Social Sciences Experiments

Cause and effect

- Life: daily questions about cause and effect
- Decisions: in part, need cause and effect thinking
- Does eat more vegetables makes you healthier?
- Does donate for a party gives you access to policymaking?
- Do smaller classrooms improve education?
- Do give monetary rewards for parents improve school attendance?
- Can you give an example?

Cause and effect

- How can we answer to those questions?
- What methods give the right answers?
- What methods do **not** give the right answer?
- We will study these things in this course.

Cause and effect: example

- Does police violence against protestors increase population support for their cause?
- Maybe firing against protestors will drive them home?
- Evidence: police suppression of black-blocks in Sao Paulo.
- Maybe it could galvanize their popular support?
- Evidence: 2013 mass protests in Brazil.
- How should the police respond?

Cause and effect: example

- Medicine: aorthic arrhythmia
- Theory: arrhythmia is a precursor to hearth attack.
- Three drugs developed to stop arrhythmia.
- Guess what?
- Big clinical trial: all drugs failed. Two of them increased the chance of hearth attack.

Cause and effect: example

- Medicine: aorthic arrhythmia
- All the theory supported the claim that arrthymia caused heart attacks
- But by correlations!
- Correlation: many people with arrthymia have hearth attack
- Does it make it a cause of hearth attack?

Cause and effect: example

- Does take a prep-course increase your scores?
- Many would think it is true.
- But what about the fact that many people taking a prep-test are motivated to get best scores?
- Mayor in the US: every time our team wins, we party.
- Would you prescribe partying as a way for your team to win?

Cause and effect

- Although this sounds ridiculous, that what we do in most of our research...
- We usually say to people: in order to pass the exam, you have to take the prep test.
- However, what are the *unobserved* factors affecting this prescription?
- *Unobserved*: things that we cannot measure: motivation, willingness to work hard, family support, etc
- Most of the research carried out claims that *we just need to control for X*

Cause and effect

- I hate to be the bearer of bad news, but there are things we cannot control for!
- How many people said to you past year:
  - You just need to control for X?
  - Your measures will improve if you look at Y as a source of heterogeneity?
  - Why don't you split your sample in Z?
- There is, and there will always be, unobservable characteristics when doing observational research.

Cause and effect

- Is there a way out of this mess?
- What if you could be sure to say: I don't need to do that because my dataset is credibly not affected by this variable?

Cause and effect

- Is there a way out of this mess?
- What if you could be sure to say: I don't need to do that because my dataset is credibly not affected by this variable? - Good news: **experiments**!

Experiments

- Experiments: assign units to treatment or control.
- Treatment: get the intervention.
- Control: gets nothing.
- But how do we decide who gets what?
- One way: alternating.
- Fibiger: tested treatment of Diphthria studying people day-on-day-off in a hospital.
- Clever... but what is the problem?

Experiments

- Fisher: first person who saw this problem.
- What is the way to assign units to treatment and control in order to eliminate any systematic differences between them?
- His answer: random assignment!

Experiments

- But why random assignment?
- Physics: does not need random assignment.
- Atoms: more or less the same.
- This means that they are *interchangeable*
- But is it possible to do experiments in social sciences?

Experiments

- The answer is: in most of the cases, yes!
- Why we don't do more then?
- Because it is very hard to mimic real-life situations.

Experiments

- Example: we could show a jar for a person.
- Say that there is a ball, either red or blue.
- Then show up in the screen: *a credible source says that the ball has 70% chance to be red*.
- Finally, the person has to guess the color of the ball.
- What is this experiment testing?

Experiments

- This experiment is great, and in some sense, has the power that it isolates the key characteristics of political information.
- However, is there any other setting that could give a *more naturalistic* approach?
- That's what we do in **field experiments**.
- The name came from actual agricultural experiments.
- **Realism**: objetive of field exps. Although hard to achieve.

Experiments: realism

- Degree if realism:
  - Authenticity of the treatment
  - Participants
  - Contexts
  - Outcomes
- But what constitutes a field experiment depends on how the *field* is defined.

Field Experiments

- Field experiments: challenging to implement.
- Require:
  - Design
  - Planning
  - Pilot testing
  - Constant supervision
- Another criticism: fail to grasp big questions.
- But the field is increasing fast.

Naturally ocurring experiments

- Quasi-experiments: experiments naturally occuring or assigned by governments or institutions.
  - Vienam draft lottery.
  - Random audit in Brazilian municipalities.
  - Scheduled castes for Indian local government.
  - Size of legislature determined by population thresholds.
- But they do not envolve explicit random assignment.

Plan of this module

- Class 1 (today): experiments and definitions
- Class 2: random sampling
- Class 3: working with covariates
- Class 4: Intro to *declare design* R package
- Class 5: Non-compliance

Plan of this class

- Class 6: Non-compliance (cont'd)
- Class 7: Attrition
- Class 8: Heterogeneity
- Class 9: Mediation
- Class 10: Getting your experiment done

Plan of this class

- Books:

Main:
Gerber and Green. Field Experiments.

Stats:
Aronow and Miller. Foundations of Agnostic Statistics.

Experiments:
Morton and Williams. Experimental Political Science and the Study of Causality.

Plan of this class

Basic Stats + R:
Imai. Quantitative Social Sciences.

Declare Design:
https://declaredesign.org

Econometrics:
Angrist and Pischke. Mostly Harmless Econometrics.

Plan of this class

- The classes will be based on Gerber and Green
- But feel free to read broad!
- Please bring experiences from your own work!
- And keep an open and creative mind!
- You homerun experiment might be just a few neuronal connections away!

Causal Inference and Experimentation

Potential Outcomes

- Experiments makes things easier in terms of analysis.
- But there are some technicalities that we need to learn.
- To decide, you have to understand what can be done to solve, without violating the randomness.

Potential Outcomes

- Suppose we have an experiment to access the impact health care provision
- Suppose we have the following question: do health care interventions to decrease dengue work?
- The context is Dengue prevention.
- Teams have to go to households, finding and eliminating breeding sites of *a. aegypti*.
- And we assign streets to either to teams or not.

Potential Outcomes: definitions

- **Treatment**: receiving or not a team.
- **Outcome**: dengue fever cases in each street block.

Potential Outcomes: definitions

- Let a given street block i.
- For each street block, we have the outcomes in terms of dengue, having the treatment assigned or not.
- And here are the outcomes:

Potential Outcomes

Street	Yi_0	Yi_1	tau_i
1	15	10	**-5**
2	15	15	**0**
3	30	20	**-10**
4	15	20	**5**
5	20	10	**-10**
6	15	15	**0**
7	30	15	**-15**
Average	**20**	**15**	**-5**

- $Y_i(0)$: outcome when the unit is **not** treated
- $Y_i(1)$: outcome when the unit is treated
- And for example, unit $i = 3$ has an effect of -10.

Potential Outcomes

Street	Yi_0	Yi_1	tau_i
1	15	10	**-5**
2	15	15	**0**
3	30	20	**-10**
4	15	20	**5**
5	20	10	**-10**
6	15	15	**0**
7	30	15	**-15**
Average	**20**	**15**	**-5**

- Treatment effects:

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

Potential Outcomes

Street	**Yi_0**	**Yi_1**	**tau_i**
1	15	10	**-5**
2	15	15	**0**
3	30	20	**-10**
4	15	20	**5**
5	20	10	**-10**
6	15	15	**0**
7	30	15	**-15**
Average	**20**	**15**	**-5**

- What is the problem here?

Potential Outcomes

Street	Yi_0	Yi_1	tau_i
1	15	10	**-5**
2	15	15	**0**
3	30	20	**-10**
4	15	20	**5**
5	20	10	**-10**
6	15	15	**0**
7	30	15	**-15**
Average	**20**	**15**	**-5**

- What is the problem here?

Potential Outcomes

- Suppose that we assign the treatment d_i to the street blocks.
- $d_i \in \{0, 1\}$
- Then, we will observe $Y_i(d_i)$ for every unit i.
- The observed outcome is equal to

$$Y_i \; = \; d_i Y_i(1) - (1 - d_i) Y_i(0)$$

- It is the combination of the two possibilities.

Potential Outcomes

- And the average treatment effect (ATE) is the average of all τ_i.

$$ATE = \frac{1}{N}\sum_{i=1}^{N}\tau_i$$

- Now suppose that we draw two street blocks randomly. How many possible selections we have?

$$\binom{N}{k} = \binom{7}{2} = 21$$

- We could now compute the average for any of the random draws:

```
##  [1] 15.0 22.5 15.0 17.5 15.0 22.5 22.5 15.0 17.5
## [10] 15.0 22.5 22.5 25.0 22.5 30.0 17.5 15.0 22.5
## [19] 17.5 25.0 22.5
```

Potential Outcomes

- And the average is equal to 20.
- We will look into the average treatment effect.
- This represents the average differences in control and treatment:

$$
\begin{aligned}
E[Y_i(1) - Y_i(0)] &= E[Y_i(1)] - E[Y_i(0)] \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i(1) - \frac{1}{N}\sum_{i=1}^{N} Y_i(0) \\
&= \frac{1}{N}\sum_{i=1}^{N}[Y_i(1) - Y_i(0)] \\
&= \frac{1}{N}\sum_{i=1}^{N}[\tau_i] \\
&= ATE
\end{aligned}
$$

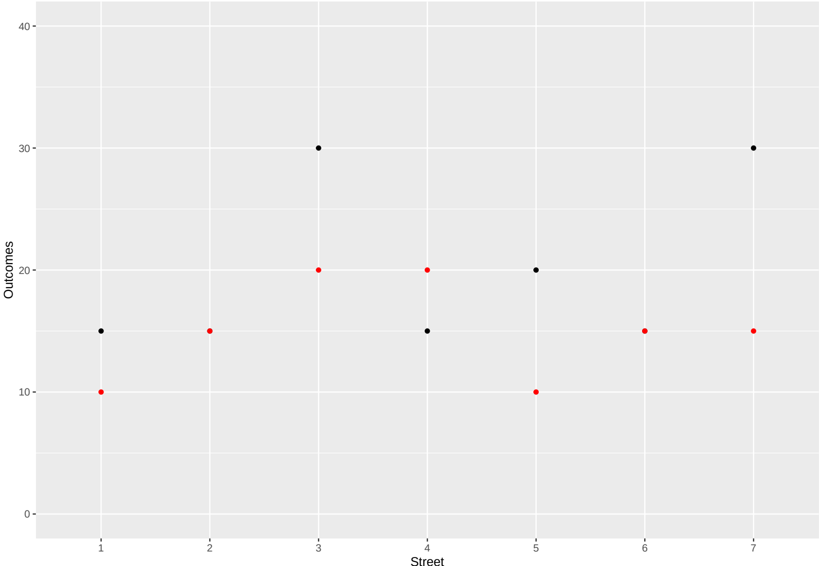
Statistics and experiments

- Causal inference is a missing data problem!

Street	**Yi_0**	**Yi_1**	**tau_i**
1	15	?	**?**
2	?	15	**?**
3	?	20	**?**
4	?	20	**?**
5	?	10	**?**
6	?	15	**?**
7	30	?	**?**
Average	**22.5**	**16**	**-6.5**

Statistics and experiments

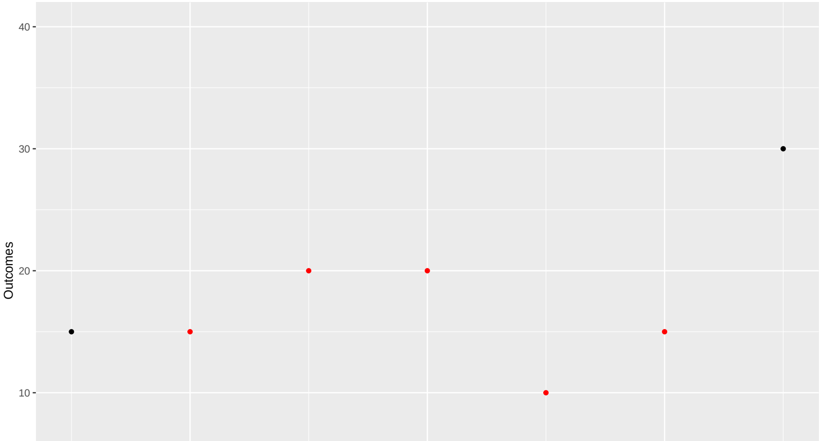
- The real effects, in case you could perfectly observe, is (red = treatment):



Statistics and experiments

- But you observe this (red = treatment):

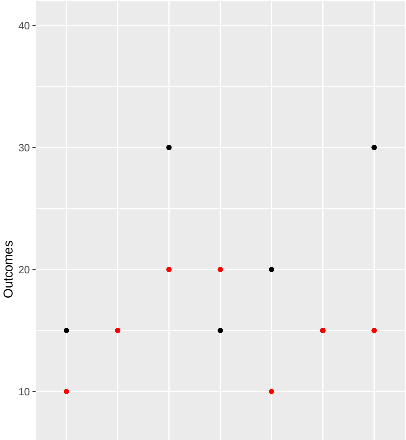
```
## Warning: Removed 5 rows containing missing values
## (geom_point).
## Warning: Removed 2 rows containing missing values
## (geom_point).
```

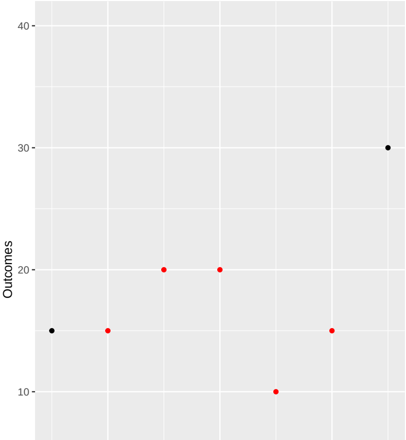


Statistics and experiments

- And side by side (red = treatment):

```
## Warning: Removed 5 rows containing missing values
## (geom_point).
## Warning: Removed 2 rows containing missing values
## (geom_point).
```





Random assignment

- Why the random assignment work?
- Let $Y_i(1)$ an unit i outcome in the treatment.
- $Y_i(0)$ an unit i outcome in the control.
- $D_i = 1$ the assignment for treatment.
- $D_i = 0$ the assignment for the control.

Random assignment

- Then, as the treatment is assigned randomly, we say that is orthogonal to the outcomes and any confounder.
- In math:

$$Y_i(0), Y_i(1), X \perp D_i$$

- And as a consequence:

$$E[Y_i(1)|D_i = 1] = E[Y_i(1)] = E[Y_i(1)|D_i = 0]$$

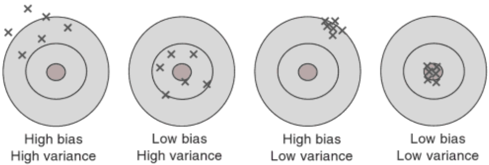
$$E[Y_i(0)|D_i = 1] = E[Y_i(0)] = E[Y_i(0)|D_i = 0]$$

- Thus:

$$ATE = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

Random assignment

- **Estimators**: what we use to find (estimate) the parameter of interest.
- Properties:



Bias Variance Decomposition. Figure 1. The bias-variance decomposition is like trying to hit the bullseye on a dartboard. Each dart is thrown after training our "dart-throwing" model in a slightly different manner. If the darts vary wildly, the learner is *high variance*. If they are far from the bullseye, the learner is *high bias*. The ideal is clearly to have both low bias and low variance; however this is often difficult, giving an alternative terminology as the bias-variance "dilemma" (*Dartboard analogy*, Moore & McCabe (2002))

Figure 1: Efficiency and unbiasedness.

- How to estimate the ATE in an unbiased way?
- If $\hat{\theta}$ is an estimator of θ, then unbiasedness mean $E[\hat{\theta}] = \theta$.

Random assignment

- Suppose we assign the treatment to m streets and the control to $N-m$ streets. Then:

$$\begin{aligned} E\left(\frac{\sum_1^m Y_i}{m} - \frac{\sum_{i=m+1}^{N-m} Y_i}{N-m}\right) &= E\left(\frac{\sum_1^m Y_i}{m}\right) - E\left(\frac{\sum_{i=m+1}^{N-m} Y_i}{N-m}\right) \\ &= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\ &= E[Y_i(1) - Y_i(0)] = E[\tau_i] = ATE \end{aligned}$$

- This estimator is unbiased! It is called the *differences-in-means* estimator.

Random assignment

- But this still does not define what a random assignment is.
- What is random assignment?
- **Definition**: An assignment that is statistically independent to all *observed* and *unobserved* variables.
- **Complete random assignment**: first design we will use.
  - When we randomly design m units for the treatment and $N - m$ units for the control.
  - Like throw a fair coin to assign cases to treatment and control.
- We will use a tool called *DeclareDesign*. It is probably the best there is on experimental implementation, analysis and diagnostics.

Random assignment

- What happens when we don't use random assignment?
- Selection bias:

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] =$$

$$E[Y_i(1) - Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

- Note that $E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$ should be zero when the assignment is random!
- Real-life example: Instead of random assignment, we left households to decide whether they will receive a given team or another.
- And the households with more Dengue get better teams.

Random assignment: core assumptions

- Potential outcomes:
  - Each street have a dengue level when treated and not treated
  - The potential outcomes, to work, require that each potential outcome depend *solely* on whether they *itself* received the treatment.
  - *solely*: excludability assumption
  - *itself*: non-interference

Random assignment: excludability

- Potential outcomes:
  - Only two, and the only relevant causal agent is the treatment assignment.
  - We must distinguish between *treatment* (d_i) and other variables that relate with the treatment (z_i).
  - Example: $Y_i(z, d)$::
    - $Y_i(z = 1, d = 1)$: get the drug and take the drug
    - $Y_i(z = 1, d = 0)$: get the drug but throw it in the garbage
    - $Y_i(z = 0, d = 1)$: wasn't supposed to get the drug, but found it in the garbage and took it
    - $Y_i(z = 0, d = 0)$: didn't get the drug and didn't take it

Excludability:

$$Y_i(z = 1, d) = Y_i(z = 0, d)$$

Threats to excludability?

- Don't take the drug.
- Don't have good measurements in a street compared to another.
- Fail to deliver the treatment (partial implementation)
- Some people refusing or claiming the treatment.
- Medical studies: double blind because of such problems...

Random assignment: non-interference

- SUTVA: Stable Unit Treatment Value Assumption
- $Y_i(d)$: written as the value of the unit only depends on the treatment itself.
- So, regardless of what was the treatment status of other units, the treatment effect depends only on the assignment to the given unit.
- For example: what if the unit get treated because we treated all neighbor street? Interference!
- If streets are far away from each other, then ok!
- We will study spillover in the next chapters.

Summary

Summary

- **Random assignment**: treatment allocated such that all units have a known probability.
  - Treatment assignment unpredictable
- **Excludability**: Potential outcomes depend only in the treatment.
- **Non-interference**: Reflect only treatment and control for the unit, and not for others.

Desenho experimental

Experimentos

- Aleatorização
- Não-interferência
- Excludability

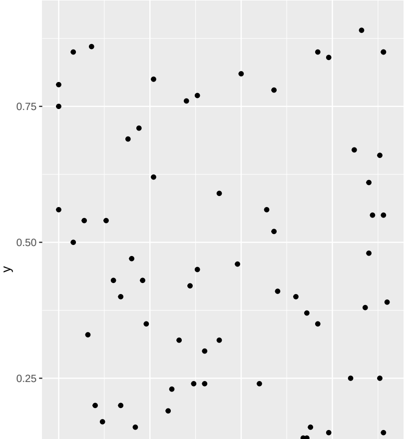
Aleatorização

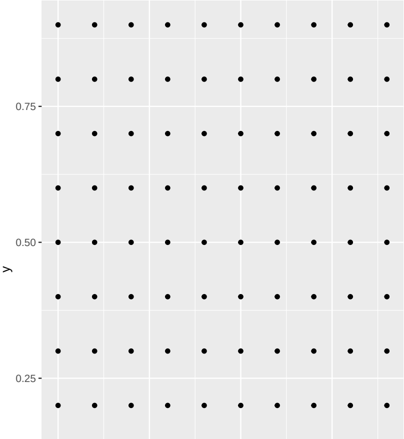
- Aleatorização é complexa.
- Computadores: produzem numeros pseudo-aleatórios.
- Aleatório: não existe padrão observável no sorteio.
- Pseudo-aleatório: existe algum padrão, mesmo que difícil de encontrar, no sorteio.

Aleatorização

- Qual dos dois você acha mais aleatório?

```
## Warning: Removed 23 rows containing missing values
## (geom_point).
## Warning: Removed 10 rows containing missing values
## (geom_point).
```





Estimação e distribuição amostral

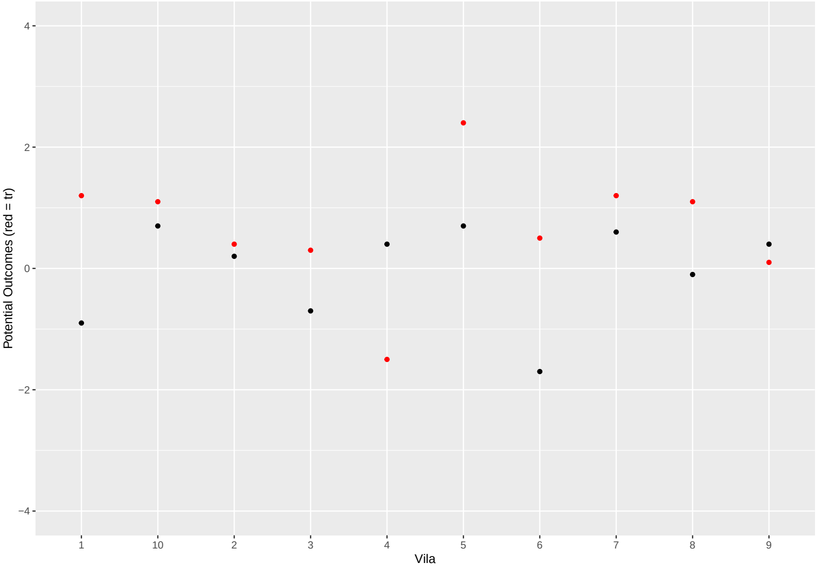
- Estatística: quantificação da incerteza.
- Queremos saber se e o quanto podemos confiar no resultado do experimento.
- Distribuição populacional: como os valores aparecem na população.
- Um exemplo de experimento *cozinhado* aqui.

Distribuição populacional

- População: 10 vilarejos
- Tratamento: Efeito de ter mulher como representante do vilarejo sobre o gasto com saneamento.
- Teoria: mulheres investem mais em saneamento que homens. Homens tendem a investir mais em estradas.
- Digamos que seja verdade...

Distribuição populacional

```
dt <- data.frame(Vila = as.character(1:10),
  Yi0 = round(rnorm(10), 1), Yi1 = 0.5+round(rnorm(10), 1))
```



Distribuição populacional

- Em tabela:

Vila	Yi0	Yi1	tau
1	-0.9	1.2	2.1
2	0.2	0.4	0.2
3	-0.7	0.3	1.0
4	0.4	-1.5	-1.9
5	0.7	2.4	1.7
6	-1.7	0.5	2.2
7	0.6	1.2	0.6
8	-0.1	1.1	1.2
9	0.4	0.1	-0.3
10	0.7	1.1	0.4

Distribuição amostral

- Para tratamentos com tamanho 5, temos 252 opções.
- Para cada uma das opções, temos os seguintes efeitos de tratamento:
- Um exemplo das combinações que podemos ter:

```
combn(10,5)
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
## [1,]    1    1    1    1    1    1    1    1    1
## [2,]    2    2    2    2    2    2    2    2    2
## [3,]    3    3    3    3    3    3    3    3    3
## [4,]    4    4    4    4    4    4    5    5    5
## [5,]    5    6    7    8    9   10    6    7    8
##      [,10] [,11] [,12] [,13] [,14] [,15] [,16]
## [1,]     1     1     1     1     1     1     1
## [2,]     2     2     2     2     2     2     2
## [3,]     3     3     3     3     3     3     3
## [4,]     5     5     6     6     6     6     7
## [5,]     9    10     7     8     9    10     8
##      [,17] [,18] [,19] [,20] [,21] [,22] [,23]
```

Calculando ATE

$$ATE = \frac{1}{N}\sum_{i=1}^{N}\tau_i$$

ATE com missing data...

- No primeiro assignment, teremos: 1, 2, 3, 4, 5. Assim, observamos:

Vila	Yi0	Yi1
1	-0.9	NA
2	0.2	NA
3	-0.7	NA
4	0.4	NA
5	0.7	NA
6	NA	0.5
7	NA	1.2
8	NA	1.1
9	NA	0.1
10	NA	1.1

- Com $E(Y_i(0)) =$ -0.06 e $E(Y_i(1)) = 0.8$.

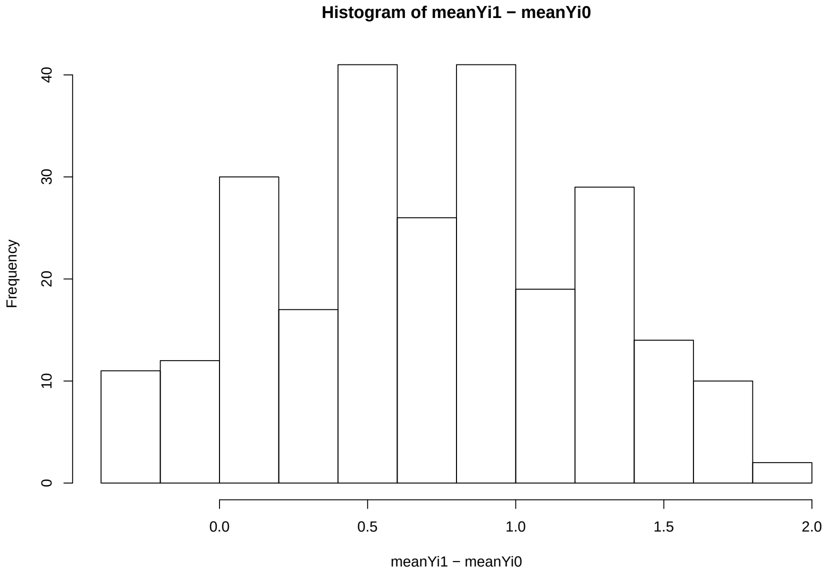
Incerteza

- Mas temos 252 possíveis combinações. Como elas se comportam?
- Podemos calcular a média para todas as combinações...

```
meanYi0 <- numeric()
meanYi1 <- numeric()
cbn <- combn(10,5)
for (i in 1:choose(10,5)) {
  meanYi0[i] <- mean(dt$Yi0[cbn[,i]])
  meanYi1[i] <- mean(dt$Yi1[11-cbn[,i]])
}
hist(meanYi1-meanYi0)
```

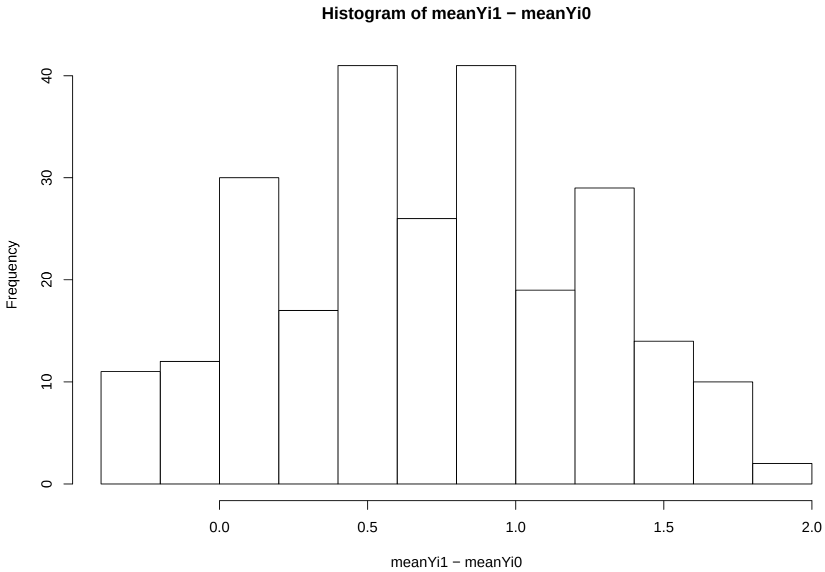
Incerteza

- E teremos:



Incerteza

- O valor real é 0.5. A média dos SATEs é 0.72! Bem próximo.



Incerteza

- Mas note algumas coisas interessantes:
  - Os valores tem variância grande!
  - Alguns valores dá tratamento maior que 1: concluiríamos que o tratamento é super efetivo
  - Alguns dão tratamento menor que 0... tratamento diminui a chance de investimentos em estradas...
  - Temos de quantificar essa incerteza!

Erro padrão

- Erro padrão é a medida do quanto de incerteza temos.
- Desvio-padrão: medida de disperção dos dados:

$$DP(X) = \sqrt{\frac{1}{N-1}(X_i - \overline{X})^2}$$

- Erro-padrão: medida de disperção do estimador:

$$EP(\widehat{ATE}) = \sqrt{\frac{1}{N-1}(ATE_i - \overline{ATE})^2}$$

Erro padrão

- Erro-padrão: formula alternativa:

$$\sqrt{\frac{1}{N-1}\left(\frac{mVar(Y_i(0))}{N-m}+\frac{(N-m)Var(Y_i(1))}{m}+2Cov(Y_i(0),Y_i(1))$$

- O que sabemos?

  1. Aumentando o N diminui o EP: aumentar a amostra!
  2. Diminuindo as variâncias diminui o EP: medir direito!
  3. Se variâncias similares, metade-metade é a melhor estratégia

Testes de hipóteses

- **Sharp null**: Não há diferenças entre tratamento em controle para todas as observações.
- **Null ATE**: Médias são iguais no tratamento e no controle
- Convencionalmente adotamos o p-valor.
- P-valor é a chance, dado que assumimos que as médias são iguais, qual a chance de detectarmos as diferenças que observamos?
- Convencionamos usar 0.05, mas isso só por conveniencia... Não tem nenhuma razão muito profunda.

Intervalos de confiança

- É um modo de ao invés de usarmos um ponto, usarmos um intervalo probabilístico.
- Definimos um nível de confiança: ex. 95%
- Calculamos a partir desse nível de confiança. . .
- Se o nível de confiança for 95%, vamos usar mais ou menos dois erros-padrão ao redor da média.
- Sammi e Aronow, 2011: Se tivermos amostras pequenas, devemos corrigir usando o fator:

$$\sqrt{\frac{N-1}{N-2}}$$

Ganhando eficiência: aleatorização por blocos

- Suponha agora que nossos 5 casos estejam em dois blocos diferentes.
- Uma maneira de aleatorizar é usar a informação que temos dos blocos.
- O que podem ser blocos?
  1. Sexo
  2. Test scores
  3. Idade
  4. Outras variáveis que temos conhecimento...

Block randomization

- Se temos dois blocos, podemos aleatorizar o tratamento dentro dos blocos.
- Suponha blocos de tamanho 4 e 6 nos nossos casos.
- Com aleatorização simples temos:

$$\binom{10}{5}$$

Que é igual a 252.

- Aleatorizando por blocos temos

$$\binom{6}{3}\binom{4}{2}$$

Que é igual a 120.

Block randomization

- E nosso ATE muda um pouco:

$$ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$$

Que é a média ponderada em cada um dos blocos.

- E o Erro Padrão do ATE muda também!

$$EP(\widehat{ATE}) = \sqrt{\sum_{1}^{J} (\frac{N_j}{N})^2 EP^2(ATE_j)}$$

Próxima aula

- Matched pair design
- Cluster randomization
- Non-compliance
- Attrition