POLI 273: Causal Inference

Professor Umberto Mignozzetti

Lecture 07 | Experimental Political Science I

Causal Inference

Experiments

- Randomization
- SUTVA
- Ignorability of the Treatment
- Good. Now what?

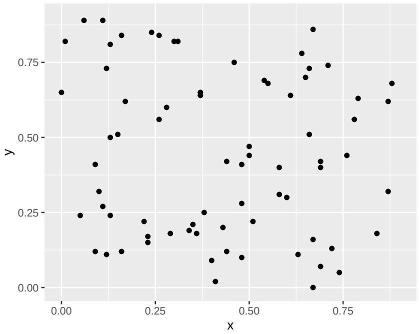
Simple Randomization

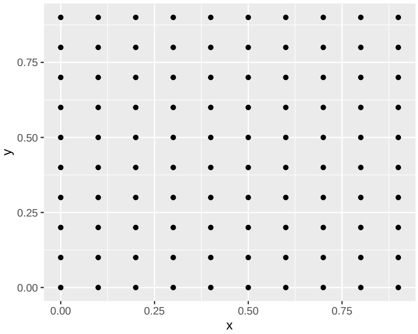
Randomization

- Randomization is complex.
- Computers: Produce pseudo-random numbers.
- Random: There are no observable patterns in the draws.
- Pseudo-random: There might be a pattern, even if hard to find, in the draw.

Randomization

- Which one do you think is *more random*?





Estimation and sampling distribution

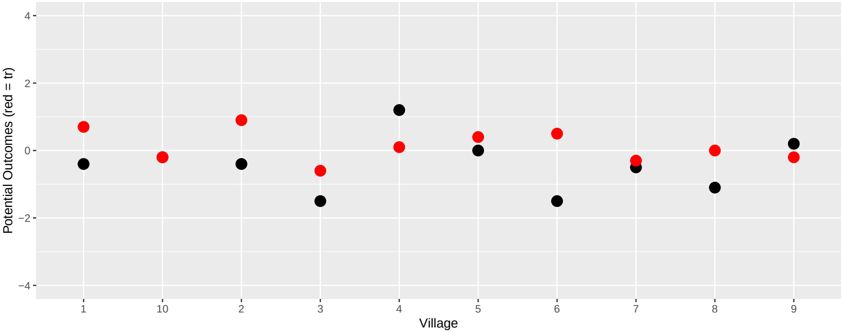
- Statistics: Quantification of uncertainty.
- We want to know **if** and **by how much** we can trust the result of the experiment.
- Population distribution: How do the values appear in the population?
- Let us see an example of a *cooked* experiment.

Population distribution

- Population: 10 villages
- Treatment: Effect of having a woman as a village representative on sanitation spending.
- Theory: Women invest more in sanitation than men.
  - Men tend to invest more in roads
- Let us say that this is true

Population distribution

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



Population distribution

- Studying the population, we would find the **PATE**: *Population Average Treatment Effect*

Village	Yi0	Yi1	tau
1	-0.4	0.7	1.1
2	-0.4	0.9	1.3
3	-1.5	-0.6	0.9
4	1.2	0.1	-1.1
5	0.0	0.4	0.4
6	-1.5	0.5	2.0
7	-0.5	-0.3	0.2
8	-1.1	0.0	1.1
9	0.2	-0.2	-0.4
10	-0.2	-0.2	0.0

- Since the ATE is: $ATE = \frac{1}{N}\sum_{i=1}^{N}\tau_i$, the *theoretical* PATE for this case is 0.55.

Sampling distribution

- But we only observe a few of these outcomes: **fundamental problem of causal inference**.
- For treatments with size *5*, we have 252 possible options!

```
combn(10,5)
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
## [1,]    1    1    1    1    1    1    1    1    1     1     1     1     1     1
## [2,]    2    2    2    2    2    2    2    2    2     2     2     2     2     2
## [3,]    3    3    3    3    3    3    3    3    3     3     3     3     3     3
## [4,]    4    4    4    4    4    4    5    5    5     5     5     6     6     6
## [5,]    5    6    7    8    9   10    6    7    8     9    10     7     8     9
##      [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26]
## [1,]     1     1     1     1     1     1     1     1     1     1     1     1
## [2,]     2     2     2     2     2     2     2     2     2     2     2     2
## [3,]     3     3     3     3     3     3     3     4     4     4     4     4
## [4,]     6     7     7     7     8     8     9     5     5     5     5     5
## [5,]    10     8     9    10     9    10    10     6     7     8     9    10
##      [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36] [,37] [,38]
## [1,]     1     1     1     1     1     1     1     1     1     1     1     1
## [2,]     2     2     2     2     2     2     2     2     2     2     2     2
## [3,]     4     4     4     4     4     4     4     4     4     4     5     5
## [4,]     6     6     6     6     7     7     7     8     8     9     6     6
## [5,]     7     8     9    10     8     9    10     9    10    10     7     8
##      [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48] [,49] [,50]
## [1,]     1     1     1     1     1     1     1     1     1     1     1     1
## [2,]     2     2     2     2     2     2     2     2     2     2     2     2
## [3,]     5     5     5     5     5     5     5     5     6     6     6     6
## [4,]     6     6     7     7     7     8     8     9     7     7     7     8
```

Computing the ATE

- In the first assignment, we have: 1, 2, 3, 4, 5. And we would observe:

Village	Yi0	Yi1
1	-0.4	NA
2	-0.4	NA
3	-1.5	NA
4	1.2	NA
5	0.0	NA
6	NA	0.5
7	NA	-0.3
8	NA	0.0
9	NA	-0.2
10	NA	-0.2

- With $E(Y_i(0)) =$ -0.22 and $E(Y_i(1)) =$ -0.04.
- The ATE for this case is 0.18.

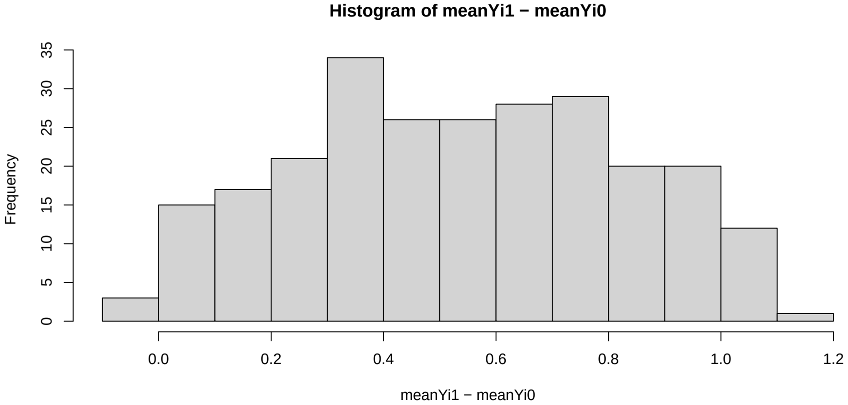
Uncertainty

- But we have 252 possible combinations. How do they behave?
- To find that out, we may compute the average for all combinations:

```
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)
```

Uncertainty in ATE measurement

- The theoretical PATE is 0.55. The mean of the ATEs is 0.55.



Uncertainty in ATE measurement

- But note some interesting things:
  1. The values have a large variance!
  2. Some values give treatment greater than one.
  - We would conclude that the treatment is **super effective**.
  3. Some give treatment effects are negative.
  - Having women decreases the chance of investments in sanitation?
- We need to think seriously about the uncertainty here.

Standard Error

- Standard error is the measure of how much uncertainty we have.
- Standard error in the ATE world is defined as:

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

- An alternative formula, that you will derive in PS02, is:

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

Standard Error

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

1. Increasing N decreases the standard-error.
   - Practical advice: Increases your sample size whenever possible.
2. Decreasing the Variances decrease the standard-errors.
   - Practical advice: Try to measure things precisely.
3. If the variances are similar, half-half is the best strategy.
   - Practical advice: You need a good reason to deviate from this benchmark.

Hypothesis Testing

- **Sharp null**: No differences in treatment and control for all units.
- **Null ATE**: Treatment and control averages differ from zero.
- p-values: We use p-values in here (sorry, Bayesian folks).
- (now for Bayesian folks): Confidence Intervals:
  - Easy to compute from this standard error formulation
  - Quick note (from Sammi e Aronow, 2011): If you have a small sample, correct the SE by the factor $\sqrt{\frac{N-1}{N-2}}$.

Block Randomization

Efficiency Gains

- Suppose now that our five cases are in two different blocks
- A way to randomize there would be to use this information to our advantage.

Block randomization

- What could be considered *blocks*?
  1. Gender
  2. Test scores
  3. Ages
  4. Other variables

Block randomization

- If we have two blocks, then we can randomize *within* these blocks.
- Assume we have blocks of size four and six.
- With simple randomization we have $\binom{10}{5}$ options, or 252 options.
- With block randomization, we would have $\binom{6}{3}\binom{4}{2}$, which is equal to 120 possible blocks.

Block randomization

- And our ATE changes slightly:

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

- The standard error also changes:

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

Weighted average/variance within blocks

Block Randomization

- Why to use block randomization?
- In some (most, really) situations, it may increase the efficiency of our estimates.
- Next class, I will bring some simulations for you to check this happening!
- Now, to fix the content, let us go to the R code for the class.

Next Class

- Matched pair design
- Population x Sample ATEs
- Block Randomization (how to do it)
- Cluster Randomization
- Block + Cluster Randomization
- *Diseases* that affect Experiments
- *Declare Design*

Questions?

See you next class!