Lecture 15: Analyzing Randomized Experiments

Prof. Esther Duflo

14.310x

Analyzing Completely Randomized Experiments

- Analyzing RCT: The conventional approach
- The Fisher exact test
- Power Calculations

The average treatment effect

- We know that, with a RCT, $E[Y^{obs}|Wi = 1] - E[Y^{obs}|Wi = 0]$ is the average treatment effect.
- How can we:
    - Find a good estimator
    - Get an estimate of the standard error of this estimator
    - Test whether it is zero
- Suppose we have a completely randomized experiment, with N_t treatment unit, and N_c control units
- What would seems a reasonable estimator for the object of interest?

Estimating treatment effect and their standard deviation

- The difference in sample average $\hat{\tau} = \frac{1}{N_t}\sum_{i:W_i=1} Y_i^{obs} - \frac{1}{N_c}\sum_{i:W_i=0} Y_i^{obs} = \overline{Y_t^{obs}} - \overline{Y_c^{obs}}$ is unbiased estimate of the treatment effect.
- The variance of a difference of two statistically independent variable is the sum of their variance, thus the variance of this estimator is $V(\hat{\tau}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$
- To estimate the variance $\widehat{V(\hat{\tau})}$ replace S_c^2 and S_t^2 by their sample counterpart:
  1. $s_c^2 = \frac{1}{N_c - 1}\sum_{i:W_i=0}(Y_i(0) - \overline{Y_c^{obs}})^2$

Estimating treatment effect and their standard deviation

- The difference in sample average $\hat{\tau} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs} - \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{obs} = \overline{Y_t^{obs}} - \overline{Y_c^{obs}}$ is unbiased estimate of the treatment effect.
- The variance of a difference of two statistically independent variable is the sum of their variance, thus the variance of this estimator is $V(\hat{\tau}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$
- To estimate the variance $\widehat{V(\hat{\tau})}$ replace S_c^2 and S_t^2 by their sample counterpart:
  1. $s_c^2 = \frac{1}{N_c-1} \sum_{i:W_i=0}(Y_i(0) - \overline{Y_c^{obs}})^2$
  2. $s_t^2 = \frac{1}{N_t-1} \sum_{i:W_i=1}(Y_i(1) - \overline{Y_t^{obs}})^2$

Estimating treatment effect and their standard deviation

- The difference in sample average $\hat{\tau} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs} - \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{obs} = \overline{Y_t^{obs}} - \overline{Y_c^{obs}}$ is unbiased estimate of the treatment effect.
- The variance of a difference of two statistically independent variable is the sum of their variance, thus the variance of this estimator is $V(\hat{\tau}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}$
- To estimate the variance $\widehat{V(\hat{\tau})}$ replace S_c^2 and S_t^2 by their sample counterpart:
    1. $s_c^2 = \frac{1}{N_c-1} \sum_{i:W_i=0}(Y_i(0) - \overline{Y_c^{obs}})^2$
    2. $s_t^2 = \frac{1}{N_t-1} \sum_{i:W_i=1}(Y_i(1) - \overline{Y_t^{obs}})^2$

Confidence intervals

- Recall our prior definition of a confidence interval: We want to find function of the random sample A and B such that $P(A(X_1 \dots X_N) < \theta < B(X_1 \dots X_N)) > 1 - \alpha$
- All we have to do is apply the lessons from the lecture on confidence intervals: we know that the ratio of the difference and the estimated standard error will follow a t distribution, so: $CI^{\tau}_{1-\alpha} = (\hat{\tau} - t_{crit} * \sqrt{(}\widehat{V}), \hat{\tau} + t_{crit} * \sqrt{(}\widehat{V}))$
- with small samples take t_{crit} from a table of t-distribution for the relevant α (as we saw in the CI lecture), with $N_T + N_C - 1$ degrees of freedom.
- with larger samples, we can use the normal approximation and take the critical value from the standard normal tables, e.g. 1.645 for $\alpha = 0.1$, and 1.96 for $\alpha = 0.05$.

Hypothesis testing

Let's start with a standard hypothesis (0 versus non zero):

$$H_o : \frac{1}{N}\sum_{i=1}^{N} Y(1) - Y(0) = 0$$

$$H_a : \frac{1}{N}\sum_{i=1}^{N} Y(1) - Y(0) \neq 0$$

Natural test statistics (following our discussion last week):
$t = \frac{\overline{Y_t^{obs}} - \overline{Y_c^{obs}}}{\sqrt{\hat{V}}}$ Follows a t distribution with $N - 1$ degrees of freedom, or with N large enough, a normal distribution.
Associated p value for two sided test : $2 * (1 - \Phi(t))$ [for the normal approximation]

Oregon Health Insurance Experiment: An example

- President Trump: "Nobody knew health care could be so complicated".

Oregon Health Insurance Experiment: An example

- President Trump: "Nobody knew health care could be so complicated".
- What are the causal effects of Affordable Care Act?
- We are lucky to have a unique experiment that tells us a lot about it.
- Oregon wanted to expand Medicaid before ACA but did not have enough money to do it: they decided to do a lottery.
- Amy Finkelstein led a team of researchers that conducted a study to follow outcomes of winners and losers of the lottery.

Some Results

TABLE 1.6

OHP effects on health indicators and financial health

	Oregon		Portland area	
Outcome	Control mean (1)	Treatment effect (2)	Control mean (3)	Treatment effect (4)
A. Health indicators				
Health is good	.548	.039 (.008)		
Physical health index			45.5	.29 (.21)
Mental health index			44.4	.47 (.24)
Cholesterol			204	.53 (.69)
Systolic blood pressure (mm Hg)			119	−.13 (.30)
B. Financial health				
Medical expenditures >30% of income			.055	−.011 (.005)
Any medical debt?			.568	−.032 (.010)
Sample size	23,741		12,229	

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on health indicators and financial health. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.

Let us spend some time with this table

- Let's compute a 95% confidence interval for the effect of insurance on the "health is good" variable.

Let us spend some time with this table

- Let's compute a 95% confidence interval for the effect of insurance on the "health is good" variable.
- (0.039-0.008*1.96; 0.039+0.008*1.96)

Let us spend some time with this table

- Let's compute a 95% confidence interval for the effect of insurance on the "health is good" variable.
- (0.039-0.008*1.96; 0.039+0.008*1.96)
- Is the hypothesis that cholesterol levels went down in Portland rejected at the 10% level ?

Let us spend some time with this table

- Let's compute a 95% confidence interval for the effect of insurance on the "health is good" variable.
- (0.039-0.008*1.96; 0.039+0.008*1.96)
- Is the hypothesis that cholesterol levels went down in Portland rejected at the 10% level ?
- No. $0.53/0.69 \leq 1.645$ (critical value for 10% level using the Normal distribution)
- What is the average physical health index in the treatment group in Portland?

Interlude: do RCT really matter

https://www.easy-lms.com/the-impact-of-extending-medicaid-coverage/course-4820

Another view of uncertainty

- With Fisher we are taking a slight detour from the statistics we have seen so far.
- We are now not going to assume that the uncertainty in our data comes from the fact that we have a *sample* drawn from an population.
- If we have the entire population, where is uncertainty coming from? Do we even need confidence intervals?

Another view of uncertainty

- With Fisher we are taking a slight detour from the statistics we have seen so far.
- We are now not going to assume that the uncertainty in our data comes from the fact that we have a *sample* drawn from an population.
- If we have the entire population, where is uncertainty coming from? Do we even need confidence intervals?
- The uncertainty is because of the missing data: each individual is *either* treated or control, but not both. And since everybody has a different potential outcome pairs (for treated and control), for each draw that nature gives us, we would get a slightly different answer.
- With big data this is the right way to think about this: Imagine running an experiment on Facebook, or using the Swedish data: this is the relevant question.

Fisher: Can we reject that the treatment has no effect on *anyone*

- Fisher was interested in the sharp Null hypothesis: $H_o : Y_i(0) = Y_i(1)$ for all i
- Note that this sharp null hypothesis is very different from the hypothesis that the average treatment effect is zero.
- The sharp null allows us to determine for each unit the counterfactual under H_o.
- The beauty of it is that we can calculate, for any test statistics we are interested in, the probability of the observed value under the sharp null
- So, suppose we choose as our statistic the absolute difference in means by treatment status: $|T^{ave}(W, Y^{obs})| = |\bar{Y}_t^{obs} - \bar{Y}_c^{obs}|$

Fisher exact test

- We can calculate the probability, over the randomization distribution, of the statistic taking on a value as large, in absolute value, as the actual value given the actual treatment assigned.
- This calculation gives us the p-value for this particular null hypothesis:

$$p = Pr(|T^{ave}(W, Y^{obs})| \geq |T^{ave}(W^{obs}, Y^{obs})|)$$

Example: Cough and Honey

- A randomized study where children were given honey or nothing.
- Main outcome: cough severity the night after the assignment (from 1 to 6)
- Imbens and Rubin (2015) use it to illustrate Fisher exact test
- First, assume we have the data for the first 6 children

the first 6 observations

Table 5.4: First Six Observations on Cough Frequency from Honey Study

Unit	Potential Outcomes		Observed Variables		
	$Y_i(0)$	$Y_i(1)$	W_i	X_i (`cfp`)	Y_i^{obs} (`cfa`)
1	?	3	1	4	3
2	?	5	1	6	5
3	?	0	1	4	0
4	4	?	0	4	4
5	0	?	0	1	0
6	1	?	0	5	1

Reference: Imbens and Rubin "Causal inference for statistics, social and biomedical sciences"

$T^{obs} = 8/3 - 5/3 = 1$

Filling out the counterfactual under the Sharp null

Table 5.5: First Six Observations from Honey Study with missing potential outcomes in brackets filled in under the null hypothesis of no effect

Unit	Potential Outcomes $Y_i(0)$	$Y_i(1)$	Observed Variables Treatment	X_i	Y_i^{obs}
1	(3)	3	1	4	3
2	(5)	5	1	6	5
3	(0)	0	1	4	0
4	4	(4)	0	4	4
5	0	(0)	0	1	0
6	1	(1)	0	5	1

Reference: Imbens and Rubin "Causal inference for statistics, social and biomedical sciences"

All the possible assignment vector, and associated statistic

W_1	W_2	W_3	W_4	W_5	W_6	levels
0	0	0	1	1	1	-1.00
0	0	1	0	1	1	-3.67
0	0	1	1	0	1	-1.00
0	0	1	1	1	0	-1.67
0	1	0	0	1	1	-0.33
0	1	0	1	0	1	2.33
0	1	0	1	1	0	1.67
0	1	1	0	0	1	-0.33
0	1	1	0	1	0	-1.00
0	1	1	1	0	0	1.67
1	0	0	0	1	1	-1.67
1	0	0	1	0	1	1.00
1	0	0	1	1	0	0.33
1	0	1	0	0	1	-1.67
1	0	1	0	1	0	-2.33
1	0	1	1	0	0	0.33
1	1	0	0	0	1	1.67
1	1	0	0	1	0	1.00
1	1	0	1	0	0	3.67
1	1	1	0	0	0	**1.00**

Reference: Imbens and Rubin "Causal inference for statistics, social and biomedical sciences"

p value?

All the possible assignment vector, and associated statistic

W_1	W_2	W_3	W_4	W_5	W_6	levels
0	0	0	1	1	1	-1.00
0	0	1	0	1	1	-3.67
0	0	1	1	0	1	-1.00
0	0	1	1	1	0	-1.67
0	1	0	0	1	1	-0.33
0	1	0	1	0	1	2.33
0	1	0	1	1	0	1.67
0	1	1	0	0	1	-0.33
0	1	1	0	1	0	-1.00
0	1	1	1	0	0	1.67
1	0	0	0	1	1	-1.67
1	0	0	1	0	1	1.00
1	0	0	1	1	0	0.33
1	0	1	0	0	1	-1.67
1	0	1	0	1	0	-2.33
1	0	1	1	0	0	0.33
1	1	0	0	0	1	1.67
1	1	0	0	1	0	1.00
1	1	0	1	0	0	3.67
1	1	1	0	0	0	**1.00**

Reference: Imbens and Rubin "Causal inference for statistics, social and biomedical sciences"
p value? $\frac{16}{20} = 0.8$

Simulation based p value

- If we have more observations we may not be able to do all the permutations (N choose k), where k is the number of treated subjects.
- Can do it as simulation: draw an assignement. Compute the statistics. repeat K times, compute the probability that the statistics is above the observed statistics.
- Example for cough and honey study (35 honey, 37 control).

Number of Simulations	p-value	(s.e.)
100	0.010	0.010
1,000	0.044	0.006
10,000	0.044	0.002
100,000	0.042	0.001
1,000,000	0.043	0.000

References

- Imbens and Rubin *Causal Inference for Statistics Social and biomedical Sciences*
- Angrist and Pishke *Mastering Metrics*