Randomized Control Trials

A. Colin Cameron
Univ. of California, Davis

These slides are part of the set of slides
A. Colin Cameron, Introduction to Causal Methods
https://cameron.econ.ucdavis.edu/causal/

March 2023

Introduction

- These slides give an introductory example of Randomized Control Trials (RCT)
  - RCTs are a method for causal inference
- In economics settings they are expensive and often difficult or impossible (for ethical reasons) to run
  - exceptions are experiments in computer labs and field experiments in development economics.
- Here we consider the cleanest case where assignment to treatment is completely random
  - similar individuals are assigned to treatment by a coin toss.
- In practice RCTs can be more complicated than this
  - then adjustment may be made using methods detailed in slides treat.pdf.

- Separately the Stata file `rct.do` implements these methods
  - using dataset AED_HEALTHINSEXP.DTA
- Data are from chapter 13.5 of A. Colin Cameron (2022) *Analysis of Economics Data: An Introduction to Econometrics* https://cameron.econ.ucdavis.edu/.
- Data originally from Aviva Aron-Dine, Liran Einav, and Amy Finkelstein (2013), "The RAND Health Insurance Experiment, Three Decades Later", *Journal of Economics Perspectives*, 27(1), pages 197-222
  - and in turn these data are from Willard G. Manning, Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz (1987), "Health insurance and the demand for medical care: evidence from a randomized experiment," *American Economic Review*, pages 251-277.

Outline

Randomized Control Trials

- A randomized control trial randomly assigns individuals to different levels of treatment.
- Example: a drug trial with individuals randomly assigned to either receiving the drug (treatment) or a placebo (control or untreated).
  - a simple assignment mechanism is to toss a coin
  - ideally the trial is a double-blind trial where neither the patient nor doctors know who received the drug and who received the placebo.
- The estimated treatment effect is simply the difference in means
  - $\bar{y}_{\text{treat}} - \bar{y}_{\text{control}}$
- Inference is based on $t = (\bar{y}_{\text{t}} - \bar{y}_{\text{c}})/\text{se}(\bar{y}_{\text{t}} - \bar{y}_{\text{c}})$
  - $\text{se}(\bar{y}_{\text{t}} - \bar{y}_{\text{c}}) = \sqrt{(s_{\text{t}}^2/n_{\text{t}}) + (s_{\text{c}}^2/n_{\text{c}})}$ if independence across individuals
    - where s_{t}^2 and s_{c}^2 are the standard deviations of y_i in the two groups.
- More generally there may be more than two levels of treatment, treatment could be continuous, treatment could be at the group level, and treated and control groups may be unbalanced.

RAND Health Insurance Experiment

- Does better health insurance increase consumption of health care?
- 1970's RAND health insurance experiment is a large social experiment
  - randomly assign different levels of health insurance to different families
  - families participate for 3-5 years.
- To ensure participants were not worse off by participating compared to their usual health insurance
  - all policies had an annual limit (MDE) after which health care was free
  - and if potentially worse off people were given a lump sum
- The experiment cost many hundreds of millions in today's dollars
  - details and results are given in Manning, Newhouse, Duan, Keeler, Leibowitz, Marquis (1987), *American Economic Review*, pp.251-277.
  - the dataset used here comes from the reanalysis by Aron-Dine, Einav and Finkelstein (2013), *Journal of Economic Perspectives*, pp.197-222.

RAND health insurance experiment (continued)

- Dataset AED_HEALTHINSEXP has 20,203 individual-year observations on 5,915 individuals in 2,205 families in the experiment for 3 years or 5 years.
- We use data for the first year of experiment and only selected variables.
  - y = total annual spending on health
  - x includes six different and mutually exclusive insurance plans ranging from 0% coinsurance (free care) to 95% coinsurance
- The coinsurance rate is the percentage of health costs paid by the individual
  - as already noted after an annual limit (the MDE) care is free in the margin (100% coinsurance).

RAND health insurance experiment (continued)

- The insurance plans are ordered by decreasing generosity (increasing coinsurance)

Variable name	Storage type	Display format	Value label	Variable label
spending	double	%9.0g		inpatient + outpatient in 2011 $
coins0	float	%9.0g		= 1 if 0% coinsurance (free) and = 0 otherwise
coins25	float	%9.0g		= 1 if 25% coinsurance and = 0 otherwise
coinsmixed	float	%9.0g		= 1 if 25%/50% mix coinsurance and = 0 otherwise
coins50	float	%9.0g		= 1 if 50% coinsurance and = 0 otherwise
coinsindiv	float	%9.0g		= 1 if individual deductible and = 0 otherwise
coins95	float	%9.0g		= 1 if 95% coinsurance and = 0 otherwise
coinsrate	float	%9.0g		coinsurance rate for plan
mde	double	%10.0g		major deductible expenditure in 1984$
spending	double	%9.0g		inpatient + outpatient in 2011 $
oop	double	%9.0g		out_of_pocket (spending not covered by insurance)

RAND health insurance experiment (continued)

- Summary statistics

Variable	Obs	Mean	Std. dev.	Min	Max
spending	5,639	1679.472	4968.068	0	175831
plan	5,639	3.310516	2.025611	1	6
year	5,639	1	0	1	1
coins0	5,639	.3321511	.4710266	0	1
coins25	5,639	.113318	.3170092	0	1
coinsmixed	5,639	.0851215	.2790871	0	1
coins50	5,639	.0663238	.2488693	0	1
coinsindiv	5,639	.2156411	.4113028	0	1
coins95	5,639	.1874446	.3903026	0	1
coinsrate	5,639	.3935136	.3508856	0	1
mde	5,639	418.4664	381.1799	0	1000
oop	5,639	228.6367	463.2244	0	4340.814

Difference in Two Means

- First consider free care insurance versus any other insurance.
  - The treatement is free insurance: coins0 $= 1$
  - The control is any other insurance: coins0 $= 0$.
- We use subscript 1 to denote the first population (coins0=1) and subscript 0 to denote the second population (e.g. badhealth=0).
- We define
  - X to be the random variable of interest (e.g. spending)
  - μ_1 and μ_0 to be the population mean of X in the two populations
  - $\bar{x}_1$ and $\bar{x}_0$ to be the sample averages in the two populations
  - σ_1 and σ_0 to be the standard deviations in the two populations
  - $s_{\bar{X}_1}$ and $s_{\bar{X}_0}$ to be the corresponding standard errors of $\bar{x}_1$ and $\bar{x}_0$.
- A 95% confidence interval for the difference $\mu_1 - \mu_0$ is then
  - estimate$\pm$1.96$\times$standard error where $se(\bar{x}_1 - \bar{x}_0) = \sqrt{s^2_{\bar{X}_1} + s^2_{\bar{X}_0}}$

$$(\bar{x}_1 - \bar{x}_0) \pm 1.96 \times se(\bar{x}_1 - \bar{x}_0).$$

- And a test $H_0 : \mu_1 = \mu_0$ against $H_0 : \mu_1 \neq \mu_0$ uses $t = (\bar{x}_1 - \bar{x}_0)/se(\bar{x}_1 - \bar{x}_0)$.

Results using a standard in two means test command

- Using a standard difference in two means test

```
. * Difference in two means: freecare (coins0=1) versus any other (coins0=0)
. ttest spending, by(coins0) unequal reverse

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       1 |   1,873     2153.57    114.6015    4959.743     1928.81     2378.33
       0 |   3,766    1443.681    80.75901    4955.998    1285.346    1602.017
---------+--------------------------------------------------------------------
Combined |   5,639    1679.472    66.15863    4968.068    1549.775    1809.168
---------+--------------------------------------------------------------------
    diff |            709.8889    140.1981                435.0165    984.7613
------------------------------------------------------------------------------
    diff = mean(1) - mean(0)                                      t =   5.0635
H0: diff = 0                     Satterthwaite's degrees of freedom =  3734.93

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

- The difference of \$709.89 is statistically significant at 5% since $p = 0.000 < 0.05$.

Results using regression on an indicator variable

- Equivalently we can regress the outcome on an intercept and an indicator variable for treatment.

```
. * Difference in two means: using OLS regression
. regress spending coins0, vce(robust)

Linear regression                               Number of obs     =      5,639
                                                F(1, 5637)        =      25.64
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0045
                                                Root MSE          =     4957.2
```

spending	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
coins0	709.8889	140.1918	5.06	0.000	435.0589	984.7188
_cons	1443.681	80.76261	17.88	0.000	1285.356	1602.007

- This gives same results
  - but can get better standard errors - here option , `vce(cluster idfamily)`
  - and can add extra variables to allow more precise estimation.

Differences in Several Means

- Now consider differences across the six mutually exclusive insurance plans.
- Regression with no intercept gives the mean spending for each coinsurance variable
- Mean spending generally drops with increasing coinsurance.

```
. * Mean spending by increasing coinsurance rate
. regress spending coins0 coins25 coinsmixed coins50 coinsindiv coins95, ///
>     vce(cluster idfamily) noconstant noheader
                              (Std. err. adjusted for 1,930 clusters in idfamily)
```

spending	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
coins0	2153.57	118.3146	18.20	0.000	1921.532	2385.608
coins25	1396.663	148.7189	9.39	0.000	1104.996	1688.33
coinsmixed	1701.874	208.1482	8.18	0.000	1293.655	2110.093
coins50	1785.845	606.6138	2.94	0.003	596.1578	2975.533
coinsindiv	1607.071	111.4304	14.42	0.000	1388.535	1825.608
coins95	1045.82	92.7984	11.27	0.000	863.8246	1227.816

Results (continued)

- OLS regression with `coins0` omitted gives difference in means compared to free care
  - average spending is less with coinsurance

```
. * Difference in mean spending compared to 0% coinsurance (free care)
. regress spending coins25 coinsmixed coins50 coinsindiv coins95, ///
>     vce(cluster idfamily) noheader
                              (Std. err. adjusted for 1,930 clusters in idfamily)
```

spending	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
coins25	-756.9073	190.0412	-3.98	0.000	-1129.615	-384.1996
coinsmixed	-451.6962	239.4243	-1.89	0.059	-921.2539	17.86156
coins50	-367.7249	618.0442	-0.59	0.552	-1579.83	844.38
coinsindiv	-546.4989	162.1506	-3.37	0.001	-864.5077	-228.4901
coins95	-1107.75	150.3658	-7.37	0.000	-1402.647	-812.8535
_cons	2153.57	118.3146	18.20	0.000	1921.532	2385.608

Results (continued)

- Test of overall significance is test of differences in means
  - $H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
  - $F(5, 1929) = 11.39$ with $p = 0.0000$ so highly statistically significant effect
  - note: standard errors cluster on family as insurance is at family level.
- We can include additional regressors.
- This is unnecessary for this RCT as it satisfied two conditions
  - there was random assignment of family to insurance plan
  - there was balance across insurance plans
    - the stratified assignment mechanism of the experiment (on age, education, , income, family size, health status) was successfully implemented
  - but it may increase the precision of estimation
  - here adding age, gender and indicators for bad health and good health makes little difference.

Further Details

- Properly conducted RCTs are regarded as the gold standard for causal inference
  - however, there are still many caveats.
- Results for an experiment need not extend to the population
  - e.g. medical trials may be restricted to certain ages, certain gender and to those healthy enough to participate.
- Extending an RCT to the population may lead to changes in the treatment effect
  - e.g. providing more generous insurance to the entire population will increase demand and hence prices of health care.
- A treatment may be statistically significant but have small effect
  - drugs are approved if they show a statistically significant effect at 5%
  - but this can arise with only a small improvement due to the drug
  - e.g. there might only be an improvement in outcome for 40% of those receiving the drug compared to 20% for those receiving the placebo.

RCTs in economics

- RCTs are difficult to run in economics, due to expense and/or ethical reasons
  - e.g. we cannot assign some to just a high school education and others to college education.
- And when they are run in economics they are often unbalanced
  - the treated and untreated groups may differ in characteristics that determine in part the outcome.
  - so use adjustment methods detailed in the slides `tr_treat.pdf`
- As a result economics relies more on so-called natural experiments or quasi-experiments with methods detailed in this set of slides
  - but as much as possible these are viewed as if they were an experiment
  - in particular specify a "counterfactual" that takes the place of a control.

References for RCTs

- A. Colin Cameron (2020), Analysis of Economics Data: An Introduction to Econometrics, chapter 13.5.
- Joshua D. Angrist and Jörn-Steffen Pischke (2015), Mastering Metrics, Princeton University Press, chapter 1.
- A. Colin Cameron and Pravin K. Trivedi (2022), Microeconometrics using Stata: Volume 2, Second Edition, Stata Press, chapter 24.3-24.4.
- Joshua D. Angrist and Jörn-Steffen Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press, chapter 2.
- Jeffrey M. Wooldridge (2010), Econometric Analysis of Cross Section and Panel Data, Second Edition, MIT Press, chapter 20.
- Guido W. Imbens and Donald B. Rubin (2015), Causal Inference in Statistics, Social, and Biomedical Sciences, Cambridge University Press, chapters 4-11.

References for RCTs (continued)

- Books by non-economists.
- Richard J. Murnane and John B. Willett (2010), Methods Matter: Improving Causal Inference in Educational and Social Science Research, Oxford University Press, chapters 4-7.
- Andrew Gelman, Jennifer Hill and Aki Vehtari (2022), Regression and Other Stories, Cambridge University Press, chapter 18.
- Specialist economics book.
- R. Glennester and K. Takavarasha (2013), Running Randomized Evaluations: A Practical Guide, Princeton University Press.