Advanced Statistical Methods for Observational Studies

LECTURE 02

class management

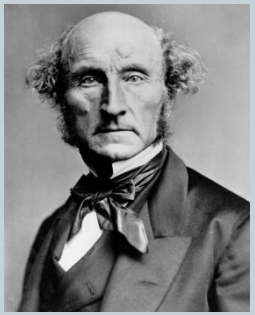
- If you want to audit then you need to let us know so we can add you to the emailing list.
- Solution to week 1 review question 5 will be posted tomorrow.
- Questions?

why randomized controlled studies produce high-quality data

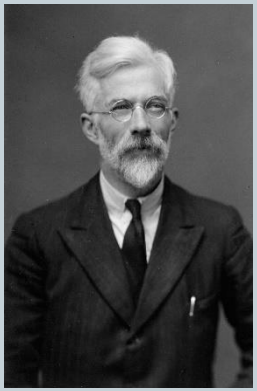
Design of Observational Studies: section 15.1 & 15.4
Holland: Statistics and Causal Inference

two approaches

- John Stuart Mill
  - Philosopher, economist, early feminist and civil servant.
  - Estimate effect through “method of differences.”



- Sir Ronald Fisher
  - Statistician and biologist.
  - Estimate effect through “a controlled & random process.”



terminology

- **Unit of observation** – the element in the study for which the intervention can be applied to or withheld from.
  - In our working example: people
  - You can imagine that we could talk about different levels of aggregation being “units of observations”: doctors who treat patients, clinics, health systems, etc.

terminology

- **Covariate** – a variable, distinct from the intervention and outcome, that can change from unit of observation to unit of observation
  - In our working example: baseline weight, gender, BMI, age, hair color, favorite color…
  - Not all covariates are equally "important." We'll revisit this notion when we discuss the concept of *confounding*.

method of difference

- In 1864, in his *System of Logic: Principles of Evidence and Methods of Scientific Investigation*, Mill proposed four methods of experimental inquiry, including the "method of difference:"

  > *If an instance in which the phenomenon … occurs and an instance in which it does not … have every circumstance save one in common … [then] the circumstance [in] which alone the two instances differ is the … cause or a necessary part of the cause (III, sec. 8)*

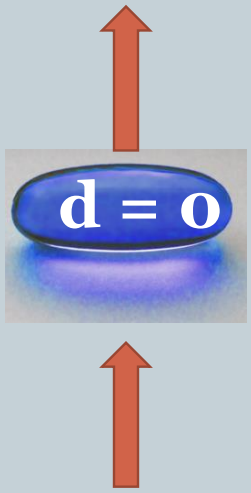
- For Mill, homogeneity and sound causal inference were closely linked: he wanted "two instances … exactly similar in all circumstances except the one" under study.

causal inference

$r_T = f(d = 1, X = x)$

x

$r_C = f(d = 0, X = x)$

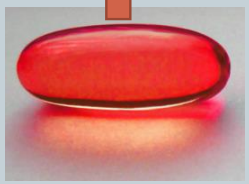


x'

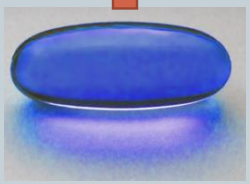
The only difference

$r_T = f(d = 1, X = \boldsymbol{x})$

$r_C = f(d = 0, X = \boldsymbol{x})$



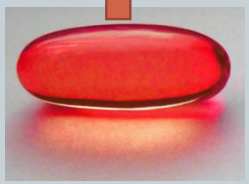
$\boldsymbol{x} =$

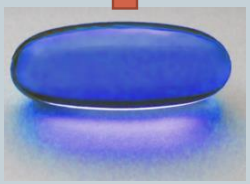


$\boldsymbol{x} =$

$\mathrm{r_T} = f(d = 1, X = x)$

$\mathrm{r_C} = f(d = 0, X = x')$





$x =$

$x' =$

terminology

- **Confounding** – *loosely speaking*, when something (usually a covariate or set of covariates) makes your estimate of the causal effect biased.
  - In our working example: baseline weight, gender, BMI, age, ~~hair color~~, ~~favorite color~~…
  - Confounding is usually thought of arising from covariates that cause variation in the outcome as well as the treatment.
  - The way I like thinking about it: If the treatment group would have had different outcomes than the control group, even if we never applied any form of intervention, then we've got confounding.

fisher: a deep insight

- Fisher's randomization
- IF we control the randomization process then we can describe, with mathematical certainty, how the data will behave.
- Armed with this understanding of data's behavior we can then make statements, with varying levels of certainty, about the state of the world.

fisher: a lady tasting tea

- In his 1935 groundbreaking book, *Design of Experiments*, he discusses an (apocryphal?) encounter he had with a lady at a gathering.
- She contended she could taste the difference between tea which had had its milk poured in first versus tea which had had milk poured in after the tea.
- Fisher thought this was hogwash and proceeded to develop a “test” of her claim.
- Interestingly, he discusses some of the reasoning that led him to this particular test.

fisher: a lady tasting tea

- In in Chapter 2 (p. 18) he wrote:

  It is not sufficient remedy to insist that "all the cups must be exactly alike" in every respect except that to be tested. For this is a totally impossible requirement in our example, and equally in all other forms of experimentation ... These are only examples of the differences probably present; it would be impossible to present an exhaustive list of such possible differences ... because [they] ... are always strictly innumerable. When any such cause is named, it is usually perceived that, by increased labor and expense, it could be largely eliminated. Too frequently it is assumed that such refinements constitute improvements to the experiment ...

confounding

- Confounding comes in two flavors:
  (i) observed confounding (the covariates are in your data set and we can probably do something), and
  (ii) unobserved confounding (you're going to have a *really* rough time…)
- We'll come back to this… but there are quite different tools based on whether or not you can justify that there is no unobserved confounding.
- Assume you have unobserved confounding, until proven otherwise.

fisher: a lady tasting tea

- His point: You will always have confounding. It will be annoying. Let's move past that.
- His proposal?
- Propose a treatment assignment process that is well-described mathematically and random
- Propose a hypothesis that will explain how the data should look in general
  - This is really important, the theory here should contain information about how the intervention interacts with the outcome,
  - The theory should guide you in which confounders are most impactful, and how to measure the outcome(s).
- Run the experiment and compare the observed data to the actual way the world worked

RCTs

- Randomized controlled trials (RCTs) are excellent
  - The “controlled” part addresses Mill’s ideas of minimizing differences at baseline
  - The “randomized” part addresses Fisher’s ideas of understanding what-else-could-have-happened

techniques for inference

Fisher's idea in practice

Design of Observational Studies: sections 2.3-2.4

testing the null of no treatment effect

- Example: Say we've got a new blood pressure medication. We recruit 20 male patients who are hypertensive. For this example, we'll concentrate on just their systolic blood pressure. We match them based on their baseline blood pressure, smoking status, age, race, and dietary habits.

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

testing the null of no treatment effect

- If we $H_0: r_C = r_T$ is true then the column assignments are merely assigned due to chance and could be reassigned without any real loss of fidelity to the way the world works.

pair	r_Ci1	r_Ti2	y_i
1	159	153	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

testing the null of no treatment effect

- If we $H_0: r_C = r_T$ is true then the column assignments are merely assigned due to chance and could be reassigned without any real loss of fidelity to the way the world works.

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

testing the null of no treatment effect

Recall that $R_{ij} = Z_{ij} r_{Tij} + (1 - Z_{ij}) r_{Cij}$.

Also note that $Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2})$.

If we $H_0: r_C = r_T$ is true…

$$Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2}) = (Z_{i1} - Z_{i2})(r_{Ci1} - r_{Ci2})$$

So if $(Z_{i1}, Z_{i2}) = (1,0)$ we get $(r_{Ci1} - r_{Ci2})$

But if $(Z_{i1}, Z_{i2}) = (0,1)$ we get $-(r_{Ci1} - r_{Ci2})$

Thus randomization flips the signs but doesn't change values.

testing the null of no treatment effect

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	133	159	-26

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	137	155	-18
6	149	140	9
7	154	148	6
8	140	145	-5
9	148	134	14
10	159	133	26

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	150	159	-9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	140	145	-5
9	148	134	14
10	159	133	26

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	140	145	-5
9	134	148	-14
10	159	133	26

testing the null of no treatment effect

- If we $H_0: r_C = r_T$ then each of those is as equally likely.
- There are 10 fair coin flips, so there are $2^{10} = 1024$ options.
- Thus each of the permutations has a chance of 1/1024.
- It's worth noting: This probability density arises by construction (coin flips) and H_0.

testing the null of no treatment effect

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

$\bar{Y} = 4.9$

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

$\bar{Y} = 6.1$

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	133	159	-26

$\bar{Y} = 0.9$

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	137	155	-18
6	149	140	9
7	154	148	6
8	140	145	-5
9	148	134	14
10	159	133	26

$\bar{Y} = 3.3$

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	150	159	-9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	140	145	-5
9	148	134	14
10	159	133	26

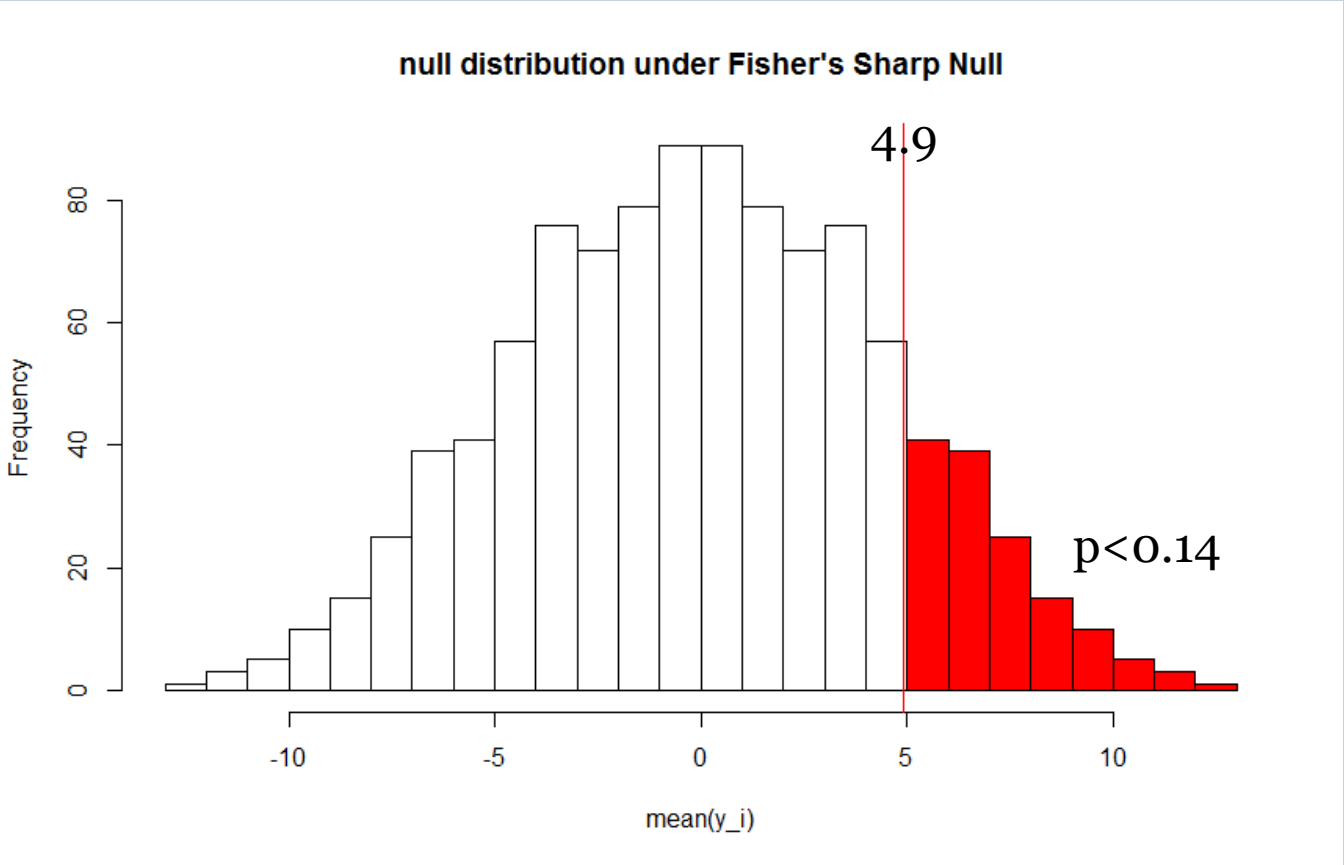
$\bar{Y} = 3.3$

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	140	145	-5
9	134	148	-14
10	159	133	26

$\bar{Y} = 1.1$

testing the null of no treatment effect

- We can then build up the null randomization distribution.



testing the null of no treatment effect

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

$s = 13.9$

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

$s = 13.32$

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	133	159	-26

$s = 14.76$

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	137	155	-18
6	149	140	9
7	154	148	6
8	140	145	-5
9	148	134	14
10	159	133	26

$s = 14.38$

pair	r_Ci1	r_Ti2	y_i
1	159	153	6
2	163	148	15
3	150	159	-9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	140	145	-5
9	148	134	14
10	159	133	26

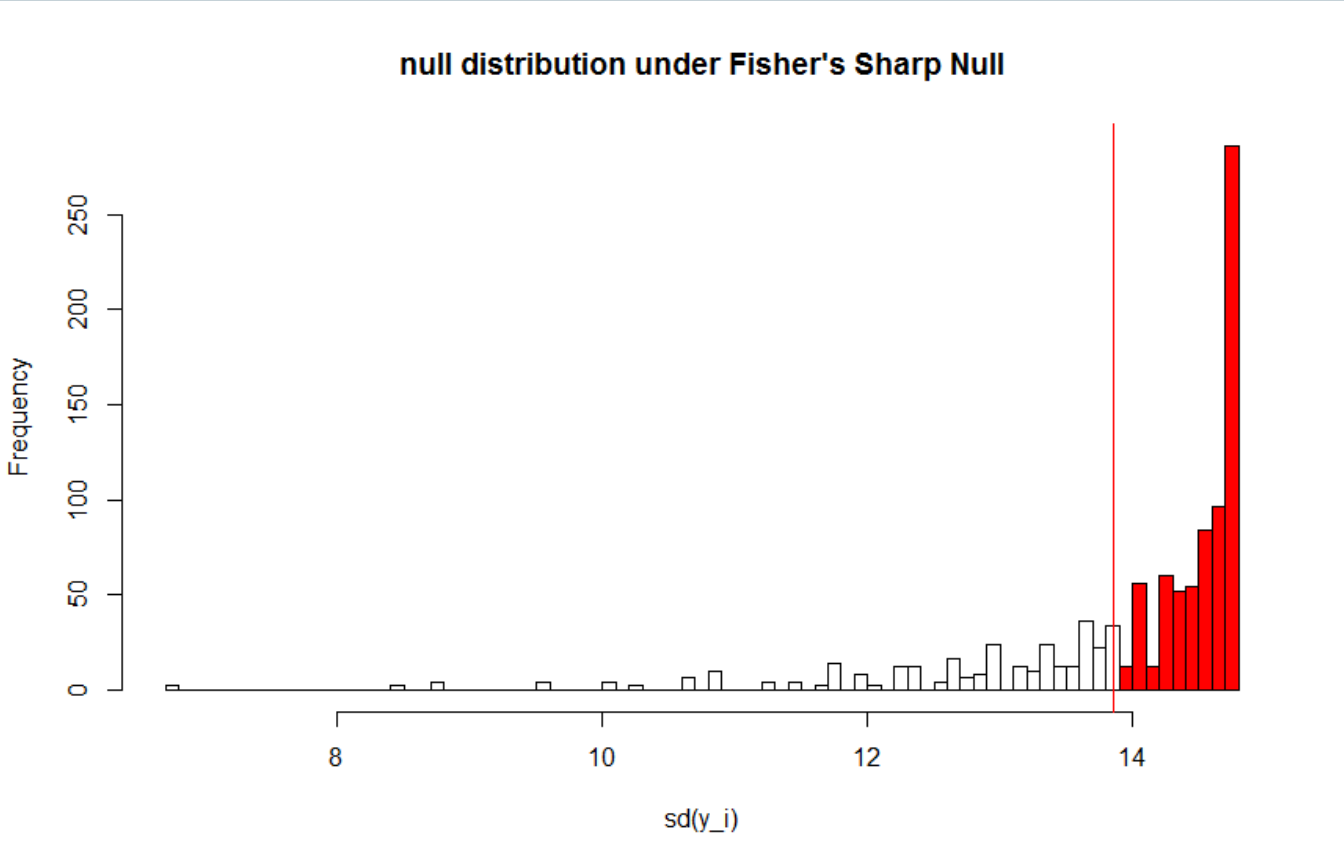
$s = 14.38$

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	140	145	-5
9	134	148	-14
10	159	133	26

$s = 14.75$

testing the null of no treatment effect

- We can then build up the null randomization distribution.



testing the null of no treatment effect

- Summary:
  - Under the null
  - Consider the randomization mechanism
  - Choose a test statistic
  - Generate the null distribution
  - Compare the observed value to the null distribution
- Generating the null distribution
  - Sample from the null distribution if really large

the Wilcoxon signed rank

- So… what's a "good" test statistic to use?
- It depends…
- A reasonable one is the Wilcoxon signed rank statistic
- How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

the Wilcoxon signed rank

How it works:

1. Take the absolute value of the difference between pairs, $|Y_i|$
2. Assign ranks to these $|Y_i|$ based on magnitude from smallest to largest, call these ranks q_i
3. The Wilcoxon signed rank statistic is the sum of the q_i where $Y_i > 0$

pair	r_Ci1	r_Ti2	y_i	abs(y_i)	q_i
1	153	159	-6	6	2.5
2	163	148	15	15	7
3	159	150	9	9	4.5
4	142	159	-17	17	8
5	155	137	18	18	9
6	140	149	-9	9	4.5
7	148	154	-6	6	2.5
8	145	140	5	5	1
9	148	134	14	14	6
10	159	133	26	26	10

4. Average ties
5. You can remove zeros

the Wilcoxon signed rank

- Takeaways
  - More robust to large deviations or corrupted data
  - Nice property is that the distribution can be calculated before the data are observed
  - Commands readily available in most packages: `wilcox.test()`

broadening the null hypothesis

- Now, instead of testing a hypothesis of "no effect," let's consider something slightly more complex:

$$H_0: r_{Cij} + \tau = r_{Tij}$$

an additive, constant effect model.

testing constant, additive effects

Recall that $R_{ij} = Z_{ij} r_{Tij} + (1 - Z_{ij}) r_{Cij}$.

If we $H_0: r_{Ci} + \tau = r_{Ti}$ is true...

$$R_{ij} = Z_{ij} r_{Tij} + \left(1 - Z_{ij}\right) r_{Cij} = r_{Cij} + Z_{ij} \tau$$

If we take a particular stance on the value of τ then we can also examine the new distribution of the data.

testing constant, additive effects

- Example: Say we've got a new blood pressure medication. We recruit 20 male patients who are hypertensive. For this example, we'll concentrate on just their systolic blood pressure. We match them based on their baseline blood pressure, smoking status, age, race, and dietary habits.

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

testing constant, additive effects

- Example: Let's suppose that $H_0: \tau = -30$.

pair	r_Ci1	r_Ti2	r_Ci2	y_i
1	153	159	189	-36
2	163	148	178	-15
3	159	150	180	-21
4	142	159	189	-47
5	155	137	167	-12
6	140	149	179	-39
7	148	154	184	-36
8	145	140	170	-25
9	148	134	164	-16
10	159	133	163	-4

```
> wilcox.test(rc,rt_tau,paired=TRUE)
Wilcoxon signed rank test with continuity correction
data: rc and rt_tau
V = 0, p-value = 0.005889
alternative hypothesis: true location shift is not equal to 0
```

the Wilcoxon signed rank

- Confidence interval
  - In general, we can test all of the values of τ_0 for $H_0: \tau = \tau_0$
  - We'd keep all of the values of τ_0 that aren't rejected and this is known as our confidence set.
  - You can imagine that in some cases that for certain hypotheses and certain test statistics that this confidence set gets really weird.
  - In practice, a lot of the hypotheses and test statistics we'll use in this class will have confidence sets that are actually intervals and that these intervals correspond with our usual confidence intervals.
  - For the Wilcoxon signed rank test of an additive effect we get a confidence interval quite easily by inverting the test.

testing constant, additive effects

- Example: Let's suppose that $H_0: r_C = r_T$.

pair	r_Ci1	r_Ti2	y_i
1	153	159	-6
2	163	148	15
3	159	150	9
4	142	159	-17
5	155	137	18
6	140	149	-9
7	148	154	-6
8	145	140	5
9	148	134	14
10	159	133	26

```
> wilcox.test(rc,rt,paired=TRUE,conf.int = TRUE)
Wilcoxon signed rank test with continuity correction
data: rc and rt
V = 37.5, p-value = 0.3323
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval: -6.000027 15.999959
```

Hodges-Lehmann point estimate of effect

- Hodges-Lehmann point estimate of effect
- The intuition is: what value of the parameter gives you data that look like they came from the null hypothesis?
- Under the null, the Wilcoxon signed rank statistic should have mean I(I+1)/4.

Hodges-Lehmann point estimate of effect

- Example: Let's suppose that $H_0: \tau = -4.5$.

$$\frac{I(I+1)}{4} = \frac{10 * 11}{4} = 27.5$$

pair	r_Ci1	r_Ti2	r_Ci2	y_i	rank
1	153	159	163.5	-10.5	5
2	163	148	152.5	10.5	5
3	159	150	154.5	4.5	2
4	142	159	163.5	-21.5	9.5
5	155	137	141.5	13.5	7.5
6	140	149	153.5	-13.5	7.5
7	148	154	158.5	-10.5	5
8	145	140	144.5	0.5	1
9	148	134	138.5	9.5	3
10	159	133	137.5	21.5	9.5

5+2+7.5+1+3+9.5=28

why randomized controlled trials do *not* produce high-quality data

venturing out of the ivory tower.

interval vs. external validity

- The inference we've been talking about have been about the unobservable causal effect $r_{Tij} - r_{Cij}$ for all of the individuals in the experiment.
- The "unknown-ness" comes from the unobserved counterfactuals. "What else could have been?"
- We're solid inside of our study, on these particular people.
- To apply the lessons learned from a study requires a bit of an extrapolation. (There's a little bit of Hume's problem of induction going on here…)
- The question of where these conclusions can be ported to is a tough question. (Active area of research: transportability)

interval vs. external validity

- Randomization is (sorta) all about internal validity.
  - Standard errors are often about how far off the counterfactuals are.
- Sampling is (sorta) all about external validity.
  - Standard errors are often about the variability from one unit to another.
- Mash them together!
  - Sample your population
  - Run an RCT on that population

beyond RCTs

- Unfortunately, RCTs can't be implemented in all situations:
  - Ethical: Does smoking cause cancer?
  - Practical: Are higher level NICUs better?
  - Expensive: Do insurance plan incentives change behavior?
  - Timeliness: Does this drug decrease depression?

naïve model for observational studies

Design of Observational Studies: section 3.1-3.3

the variation comes from somewhere

- We've built up some understanding of how to analyze data generated by a randomized controlled trial, and saw that the key feature was that the researcher directed the randomization.
- The crux of the rest of this class will be to consider methods for when the treatment assignment is not under the direction of the researcher.

the variation comes from somewhere

- The model starts with thinking about the assignment mechanism:

$$\pi_i = \Pr(Z_i = 1 | r_{Ti}, r_{Ci}, \boldsymbol{x}_i, \boldsymbol{u}_i)$$

where π_i will likely vary from person to person, $\boldsymbol{x}_i$ are the observed covariates and $\boldsymbol{u}_i$ are the unobserved covariates.

- Consider the vector of all treatment assignments, $\boldsymbol{Z}$. This can be written as:

$$\Pr(\boldsymbol{Z} = \boldsymbol{z} | \boldsymbol{r_T}, \boldsymbol{r_C}, \boldsymbol{x}, \boldsymbol{u}) = \pi_1^{z_1}(1-\pi_1)^{1-z_1} \cdots \pi_n^{z_n}(1-\pi_n)^{1-z_n}$$

motivation for matching

- Imagine we can find two subjects, k and l, such that

$$Z_k + Z_l = 1$$

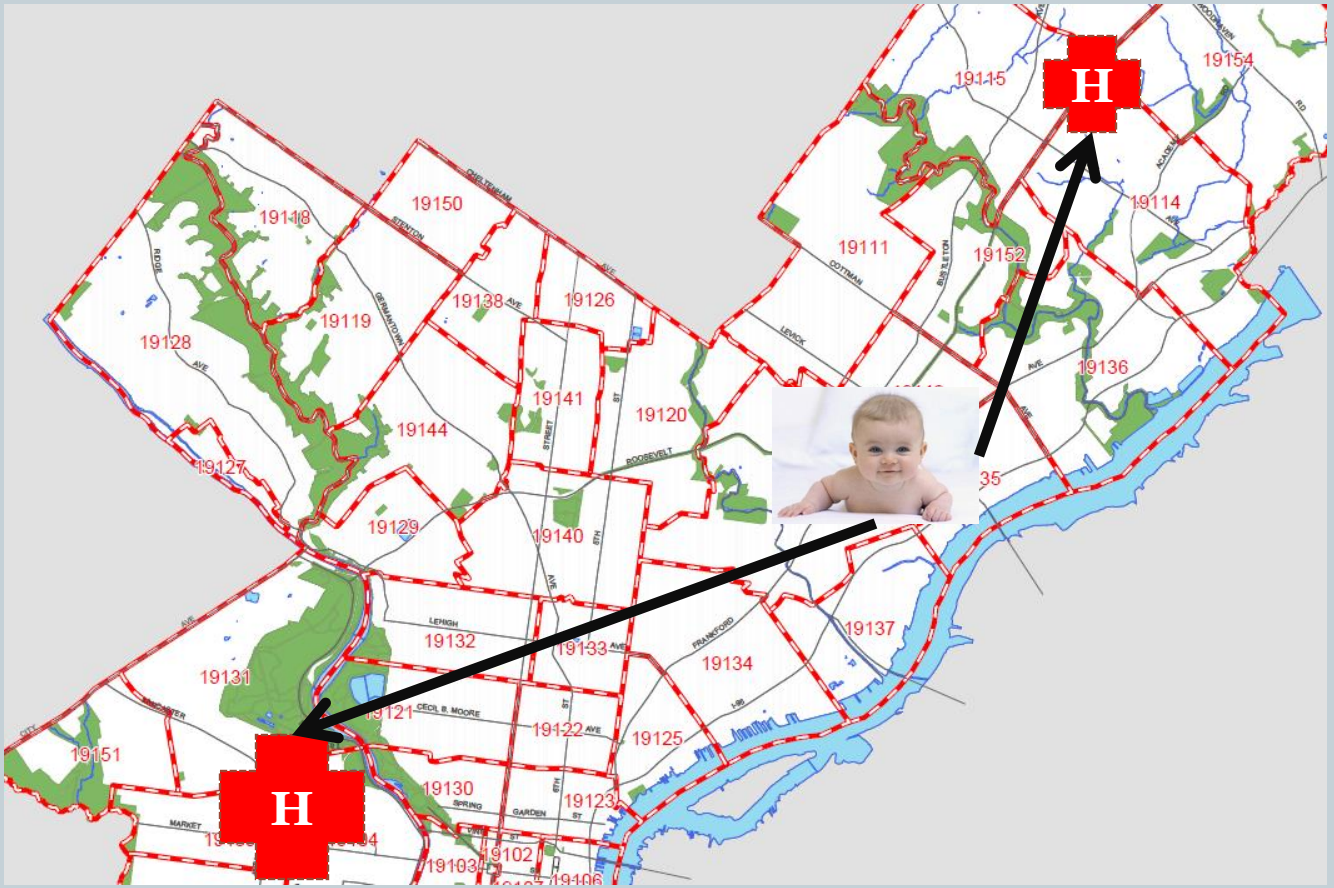
but

$$\pi_k = \pi_l$$

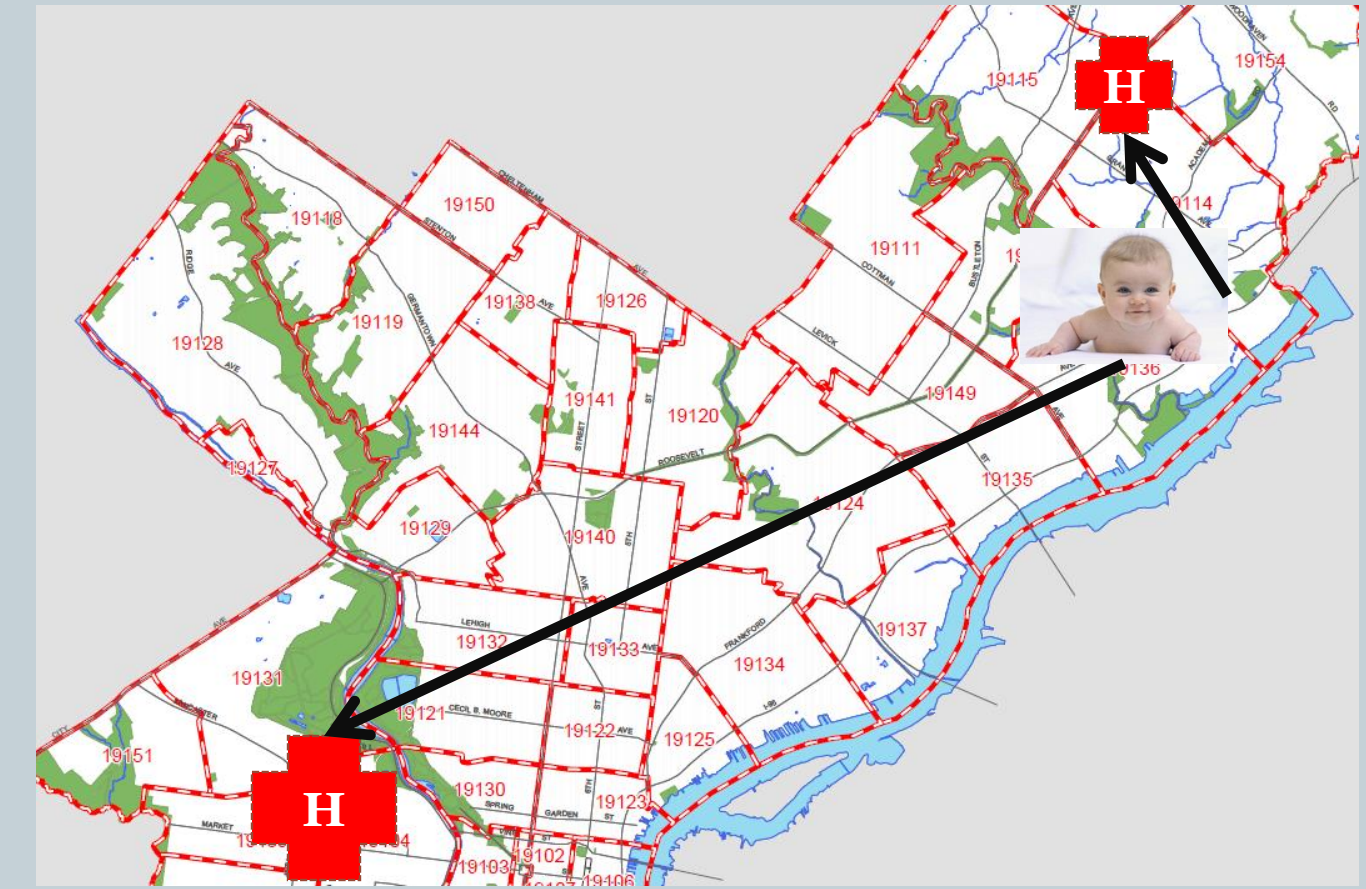
- Recall the model

$$\pi_i = \Pr(Z_i = 1 | r_{Ti}, r_{Ci}, \boldsymbol{x}_i, u_i)$$

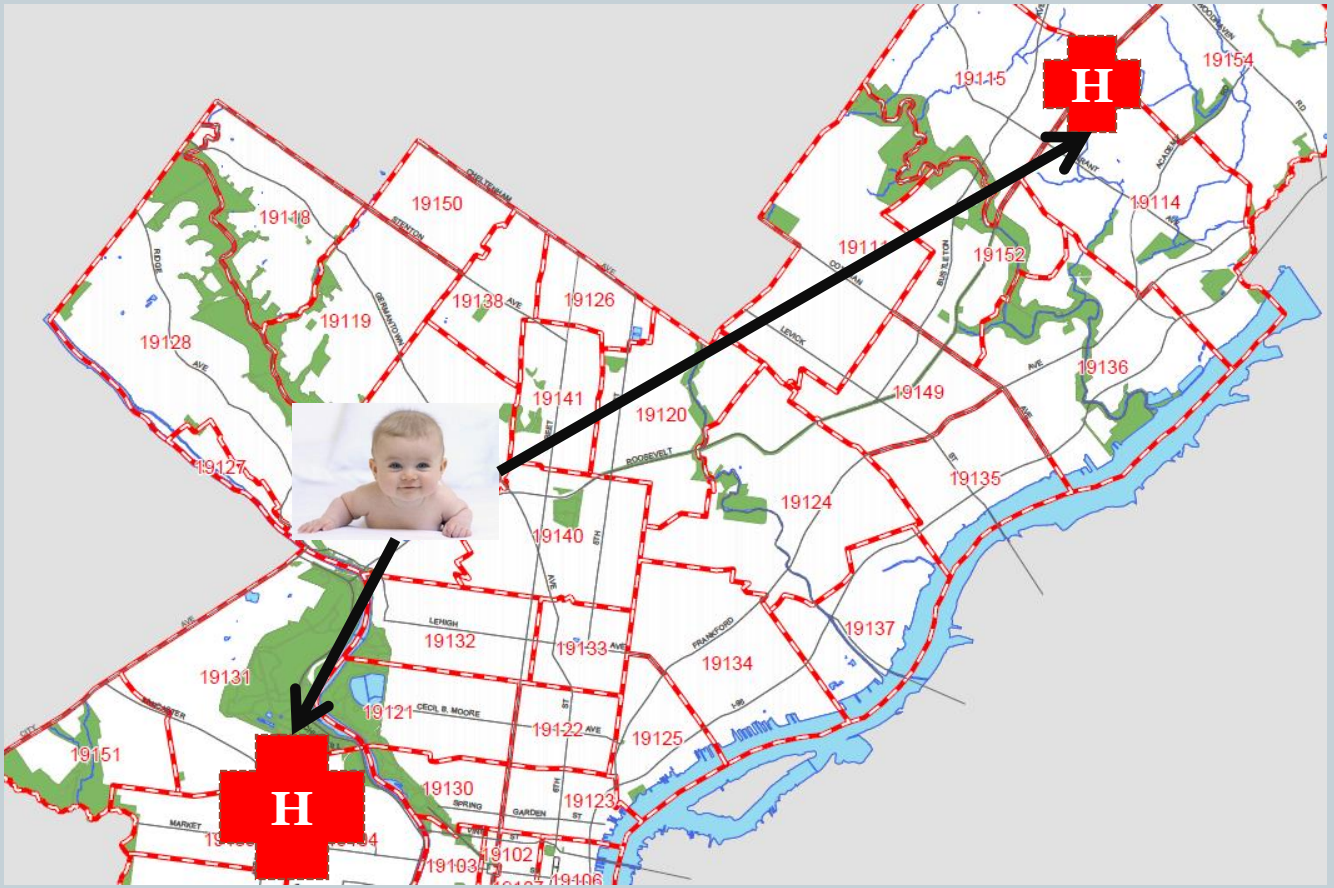
  we can only match on a subset of these.
- If we were back in the RCT framework then we could create these trivially by doing a matched pairs randomization and assigning k and l to the same pair.
- We're building a story about the randomization.



Excess Travel Time



Excess Travel Time



Excess Travel Time

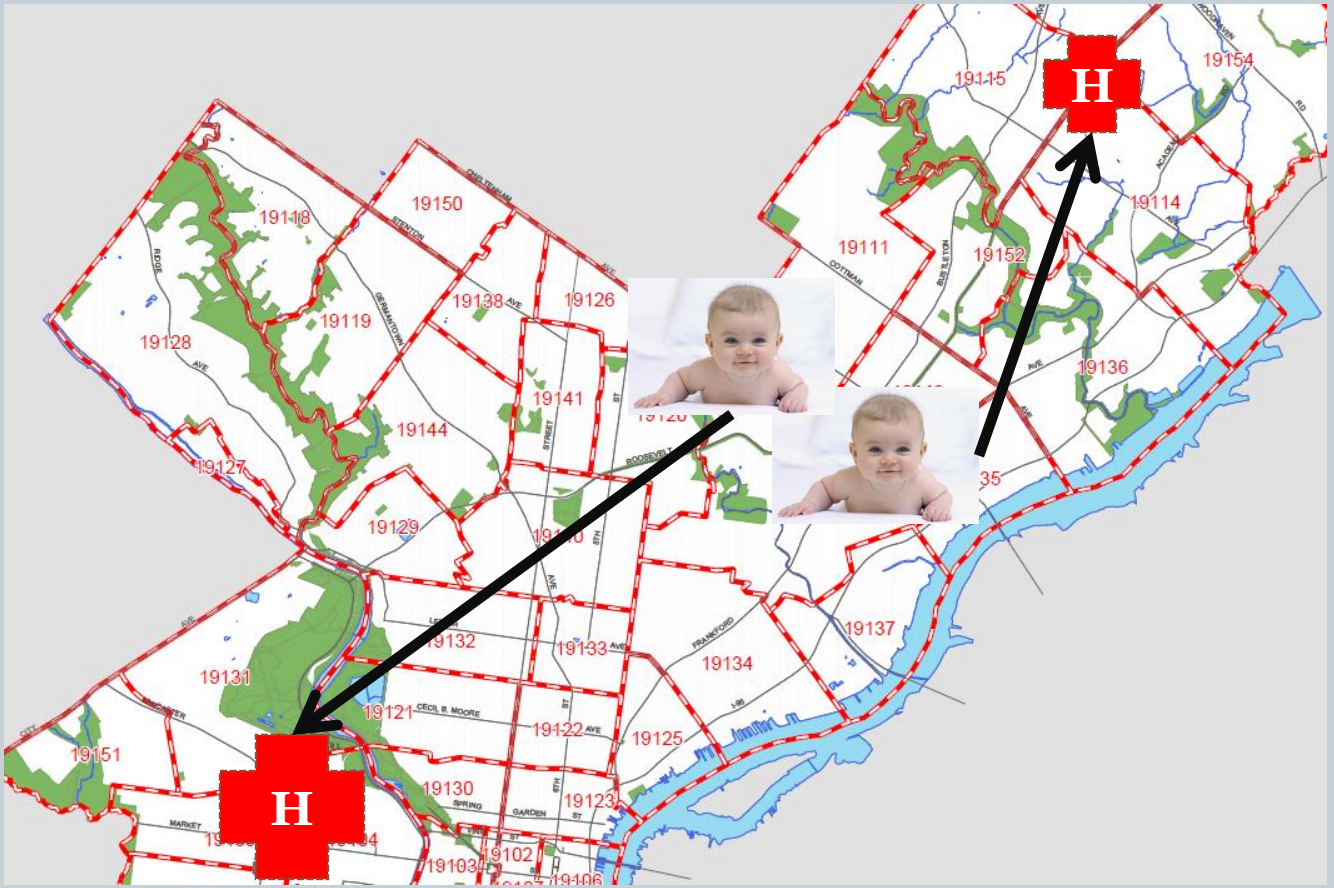
motivation for matching

Recall the model

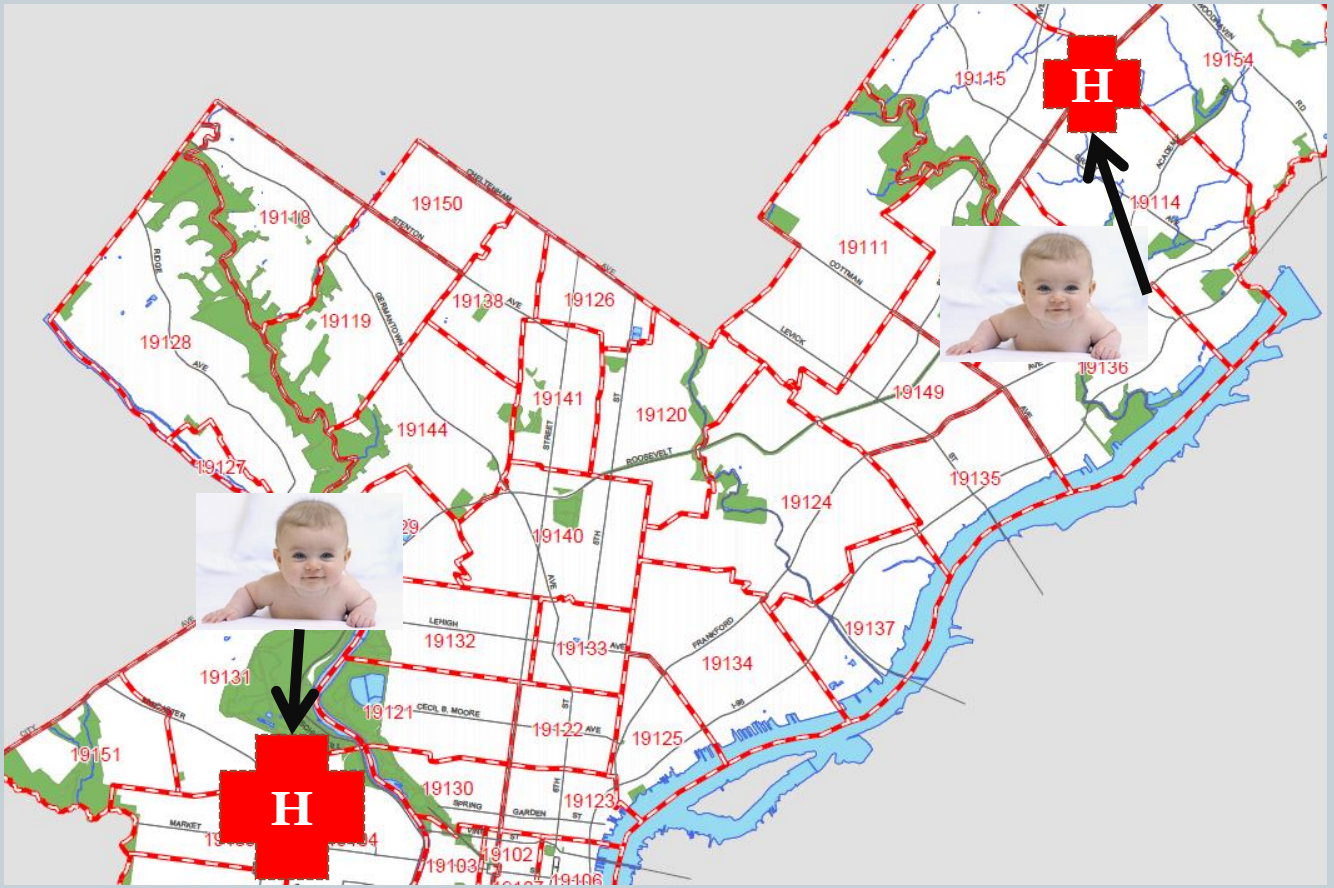
$$\pi_i = \Pr(Z_i = 1 | r_{Ti}, r_{Ci}, \boldsymbol{x}_i, u_i)$$

we can only match on a subset of these.

- r_{Ti}, r_{Ci} are whether baby lives or dies.
- Z_i is whether baby is delivered at low or high level NIC.
- $\boldsymbol{x}_i$ are the 40+ covariates of mom and baby, as well as where they live.
- u_i are all the unmeasured features that influenced the baby, mom and the process of how they ended up delivering at the hospital they did.



Selection is potentially biased!



Selection largely due to the distance!

motivation: “natural” experiments

- Perhaps the π_i behave in ways that will allow us to draw inferences in ways that are similar to experiments.
- Loosely speaking, this will tend to happen when the assignment mechanism is “haphazard” in its assignment – not using prognostically relevant covariates to sort the units into levels of the treatment.
- What happens if we can do if we do find matches such that $Z_k + Z_l = 1$ and $\pi_k = \pi_l$?

motivation: "natural" experiments

- What happens if we can do if we do find matches such that $\pi_k = \pi_l$?

$$\Pr(Z_k = z_k, Z_l = z_l | r_{Tk}, r_{Ck}, \boldsymbol{x}_k, u_k, r_{Tl}, r_{Cl}, \boldsymbol{x}_l, u_l)$$
$$= \pi_k^{z_k}(1-\pi_k)^{1-z_k}\pi_l^{z_l}(1-\pi_l)^{1-z_l}$$
$$= \pi_k^{z_k+z_l}(1-\pi_k)^{(1-z_k)+(1-z_l)}$$

- What if we design our study such that $Z_l + Z_k = 1$?

$$\Pr(Z_k = 1, Z_l = 0 | \ldots, Z_l + Z_k = 1)$$

motivation: "natural" experiments

- What happens if we can do if we do find matches such that $\pi_k = \pi_l$?

$$\Pr(Z_k = z_k, Z_l = z_l | r_{Tk}, r_{Ck}, \boldsymbol{x}_k, u_k, r_{Tl}, r_{Cl}, \boldsymbol{x}_l, u_l)$$
$$= \pi_k^{z_k + z_l}(1 - \pi_k)^{(1-z_k)+(1-z_l)}$$

- What if we design our study such that $Z_l + Z_k = 1$?

$$\Pr(Z_k = 1, Z_l = 0 | \ldots, Z_l + Z_k = 1)$$

$$= \frac{\Pr(Z_k=1, Z_l=0|\ldots)}{\Pr(Z_k=1, Z_l=0|\ldots) + \Pr(Z_k=0, Z_l=1|\ldots)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\Pr(Z_k=1, Z_l=0|\ldots) + \Pr(Z_k=0, Z_l=1|\ldots)}$$

motivation: "natural" experiments

- What if we design our study such that $Z_l + Z_k = 1$?

$$\Pr(Z_k = 1, Z_l = 0 | \dots, Z_l + Z_k = 1)$$

$$= \frac{\Pr(Z_k=1,Z_l=0|\dots)}{\Pr(Z_k=1,Z_l=0|\dots)+\Pr(Z_k=0,Z_l=1|\dots)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\Pr(Z_k=1,Z_l=0|\dots)+\Pr(Z_k=0,Z_l=1|\dots)}$$

$$= \frac{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)}}{\pi_k^{1+0}(1-\pi_k)^{(1-1)+(1-0)} + \pi_k^{0+1}(1-\pi_k)^{(1-0)+(1-1)}} = \frac{1}{2}$$

IF we can do this then we get to use the same tools developed for RCTs!

the naïve model

- Those that look alike (in our data set) are alike:

$$\pi_i = \Pr(Z_i = 1 | r_{Ti}, r_{Ci}, \boldsymbol{x}_i, u_i) = \Pr(Z_i = 1 | \boldsymbol{x}_i)$$

and

$$0 < \pi_i < 1 \text{ for all } i = 1, 2, \ldots, n$$

with

$$\Pr(\boldsymbol{Z} = \boldsymbol{z} | \boldsymbol{r_T}, \boldsymbol{r_C}, \boldsymbol{x}, \boldsymbol{u}) = \prod_{i=1}^{n} \pi_i^{z_i} (1 - \pi_i)^{1 - z_i}$$

- We could make this model true by randomly assigning, potentially using biased coins based on the observed covariates.

at the core: the propensity score

- This quantity has a name:

$$e(\boldsymbol{x}_i) = \Pr(Z_i = 1|\boldsymbol{x}_i)$$

  often, $e(\boldsymbol{x}_i)$ is referred to as the *propensity score*.
- It's the probability of assignment to treatment.
- It links the features of an individual unit to its assignment to the treatment.
- It's a description of the assignment mechanism.
- It's not magic.
- But it does have several cool features that are not immediately obvious:
  - Propensity: facilitates randomization (Fisher)
  - Score: short for "balancing score" which facilitates balance (Mill)

wishes

- If the ideal match and the naïve model were true then we could just match on the observed covariates and analyze data using traditional techniques for RCTs.
- Unfortunately, it's extraordinarily unlikely that strongly ignorable treatment assignment actually holds.
- It can be more plausible or less plausible given different situations, but it can never be proven.

fin.