Statistics II

Week 6: **Matching**

Lecture Review

Thinking about experiments

Conditional randomization: choose relevant covariates; random treatment assignment within (combinations of) covariate levels (a.k.a. **randomized block design**).

Paired randomization: as above; only two subjects per (combination of) covariate value (but multiple covariate-identical pairs allowed), one of which is randomly assigned to the treatment.

Randomized block design

- Subjects are assigned to blocks, based on gender
- Within each block, subjects are randomly assigned to treatments (placebo or vaccine)
- It is thought that men and women may react differently to this medication
- This design ensures that each treatment condition has an equal proportion of men and women
- As a result, differences between treatment conditions cannot be attributed to gender.

Gender	Treatment	
	Placebo	*Vaccine*
Male	250	250
Female	250	250

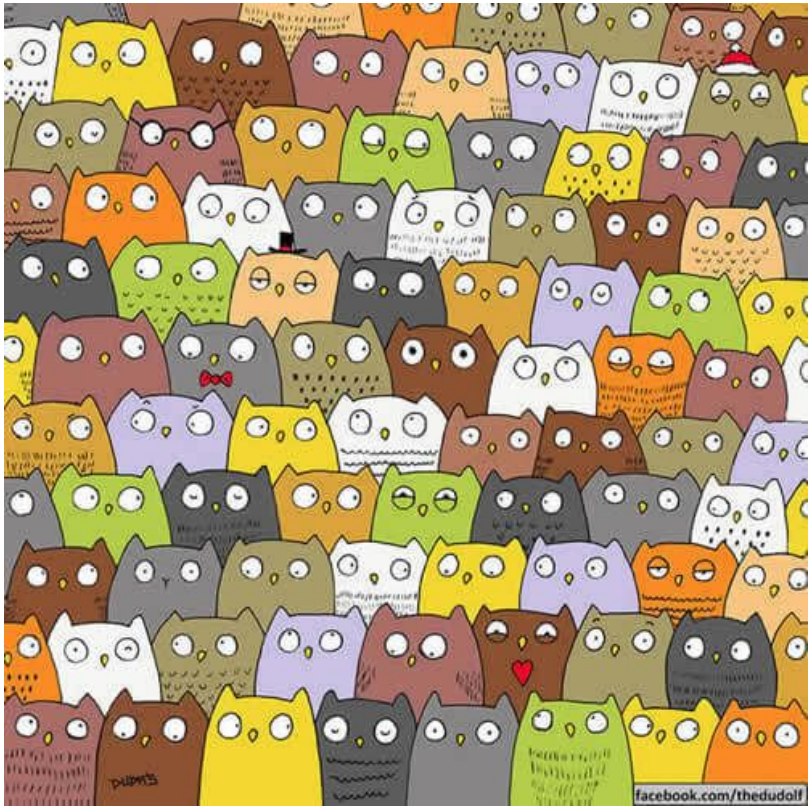
Paired randomization design

- Subjects are grouped into pairs based on some blocking variable(s)
- Within each pair, subjects are randomly assigned to different treatments
- Below, 1000 subjects are grouped into 500 matched pairs
- Each pair is matched on gender and age
- For example, Pair 1 might be two women, both age 21. Pair 2 might be two men, both age 21. Pair 3 might be two women, both age 22; and so on.

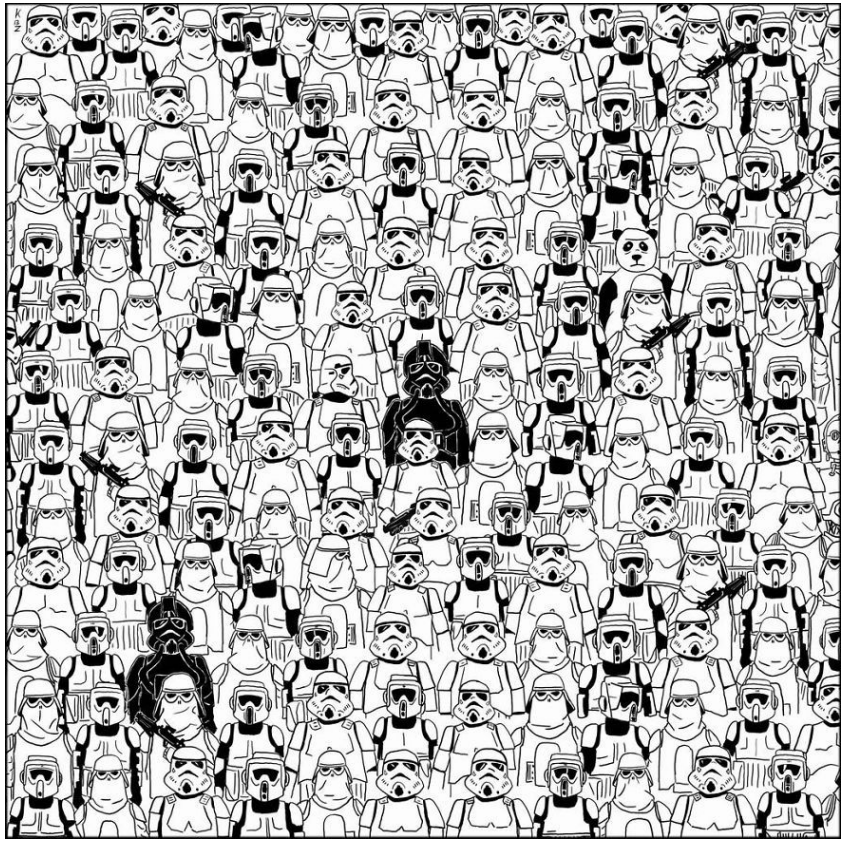
Pair	Treatment	
	Placebo	*Vaccine*
1	1	1
2	1	1
...	...	...
500	1	1

Let’s do an attention exercise

Where is the cat?



Where is the panda?



Let’s assess the effect of coffee on attention

Who had **coffee** this morning?

Identify other confounding factors for **coffee** on concentration.

Let’s come up with an estimate of the causal effect of **coffee** on concentration levels (measured in tasks completed)*

Subclassification / observational analogue

Motivation: when dealing with the problem of possible confounders for causal inference, subclassification is another way to satisfy the backdoor criterion and achieve conditional independence between treatment (D) and plausible confounders (W).

Main Idea: Compare apples to apples. Find observations that *only differ* significantly in their treatment status and estimate treatment effect among such pairs.

How? Split our sample in strata (levels) of the potential confounder, and observe the effect of treatment within those strata.

Conditioning using subclassification

Raw numbers

Internet use	Right-wing support		
	yes	no	total
yes	40	33	73
no	11	16	27

Shares

Internet use	Right-wing support
yes	0.55
no	0.41

$$NATE = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$
$$= 0.55 - 0.41 = 0.14$$

Numbers of right-wing supporters by internet use and age class

Internet use	Aged 20–40 Right-wing support		Aged 41–60 Right-wing support		Aged 60+ Right-wing support	
	yes	total	yes	total	yes	total
yes	9	18	28	40	3	15
no	1	2	7	10	3	15

Shares

yes	0.5	0.7	0.2
no	0.5	0.7	0.2

$$NATE_{overall} = NATE_{20-40} \times \frac{20}{100} + NATE_{41-60} \times \frac{50}{100} + NATE_{60+} \times \frac{30}{100} = 0.$$

We can achieve balance in the distribution of covariates between treatment and control groups by weighting differences in means within specific strata.

Matching

Through matching techniques we seek to explicitly balance the distribution of covariates between treatment and control groups.

To overcome the lack of ‘twins’ to compare treated and controlled units, we can match observations to the most plausible counterfactual available.

There are multiple ways do define what “most plausible” means. We must choose a technique for that purpose:

- Mahalanobis/nearest neighbor covariate matching
- Propensity score matching
- Coarsened exact matching

Exact Matching

1. Use theoretical and empirical knowledge to identify relevant confounder(s) (X)

2. Starting from treated subjects, select at least one match from the control group with exactly the same value(s) on *X*

3. Drop subjects off “common support” (unmatched subjects)

4. Estimate causal effect as the average difference in Y across pairs of matched subjects.

$$\hat{\delta}_{ATT} = \frac{1}{N_T} \sum_{D_i=1} (Y_i - Y_{j(i)}) = \frac{1}{N_T} \sum_{D_i=1} (Y_i - [\frac{1}{M} \sum_{D_i=1} Y_{j_m(1)})$$

If there is more than one match, you can use their average outcome as the counterfactual.

Propensity score matching

A propensity score is a measure of the **predicted probability of being in the treatment group**, given the relevant covariates (W).

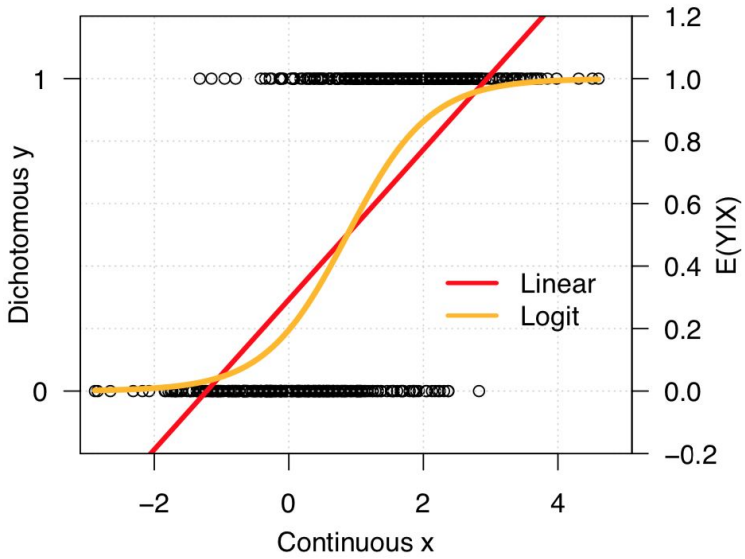
We can use propensity scores in order to match treated units with control observations that *look as if* they were treated.

This is usually modeled with logit/probit regression by which all the potential confounders are used to estimate the single value (PS).

Logit/probit regression

Similar to linear regression, except we're working with a binary categorical outcome variable.

Instead of fitting a line to the data, it fits an S-shaped curve that goes from 0 to 1. It tells you the probability of outcome based on the covariates –this is our propensity score!



Propensity score matching

- If the model for estimating the propensity score is well specified (ie. if we chose the right covariates to fulfill back-door criterion), we can control for (match on) the propensity scores and achieve conditional independence.

$$Y_0, Y_1 \perp\!\!\!\perp D | pr(W)$$

- When there are no exact matches on PS, we can define an algorithm to find the most plausible counterfactual based on PS → implies defining issues like replacement, caliper/trimming.

(Very general) steps for matching using propensity scores

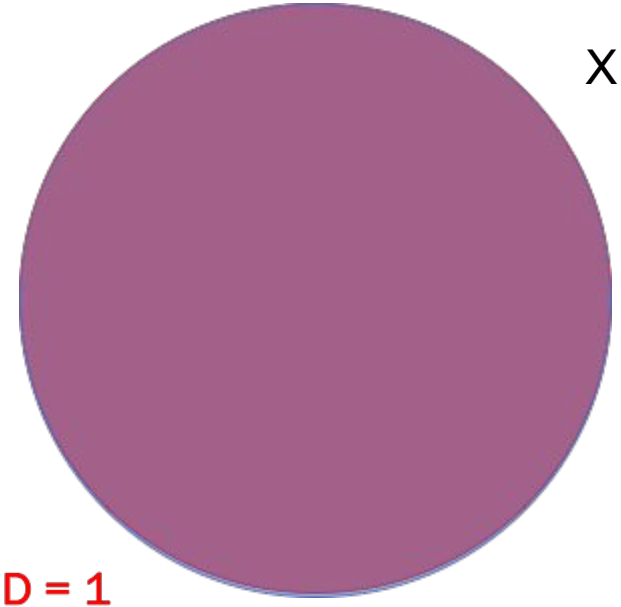
- Define the set of **potential confounders** (W) by laying out the causal graph.
- Model the probability of P(D = 1 | X), using a **logit/probit regression model.**
- Use predicted treatment probabilities as an estimate of **propensity scores**.
- Inspect PS distribution to define whether to **trim** or not. (discard observations unlikely to have a plausible match. Renounce ATE).
- **Match** subjects from treatment and control group applying an algorithm of your choice.
- Check whether your treatment (D=1) and control (D=0) groups are balanced in terms of the covariates you defined (t-tests). If not balanced, repeat.

Only then estimate treatment effect (the matching method in itself does not estimate the effect).

Common support (the curse of dimensionality)

- The more confounders we consider, the less likely it is that we find units with otherwise identical characteristics in the treatment and control groups.
- We **cannot compare all units to a 'twin'**: they lack common support.
- Without common support for all units, **we cannot estimate the ATE**.
- Knowing **which information** is missing is important. Depending on where the gaps are, we can estimate other effects.

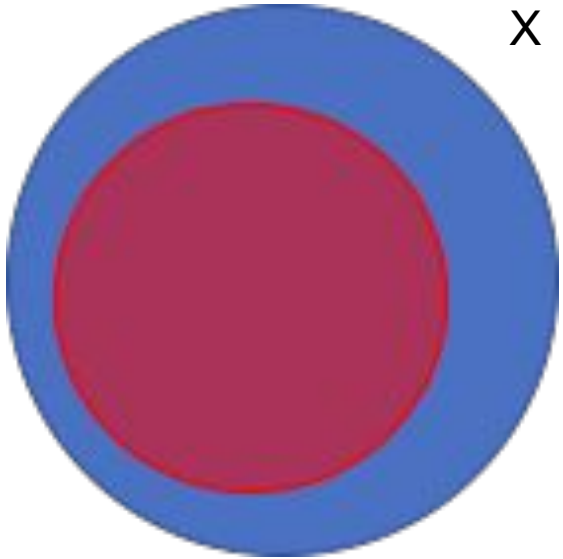
ATE



Name	Y	D	X
Jake	10	1	3
Gina	8	1	2
Terry	6	1	1
Rosa	8	0	3
Charles	6	0	2
Ray	4	0	1

In this case, we have full common support, meaning that the distributions of *X* under both treatment and control are equal. We could gather the ATE.

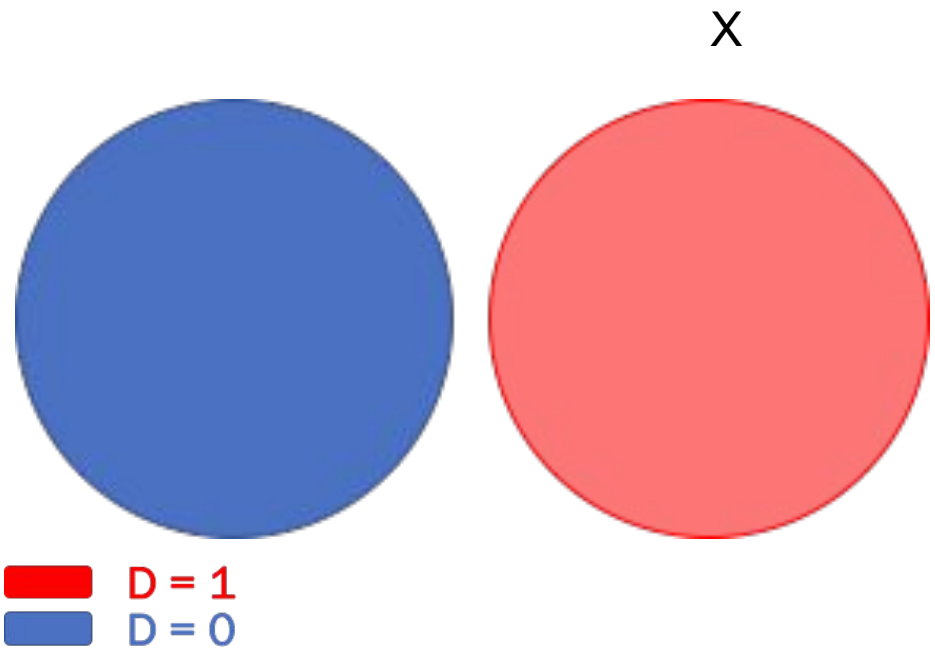
ATT



Name	Y	D	X
Jake	10	1	3
Gina	12	1	3
Terry	8	1	2
Rosa	6	0	3
Charles	3	0	2
Ray	1	0	1

In this case, all our treated units have common support in the control group. But not our controls have a “twin” in the treatment group. We can gather the Average Treatment Effect for the Treated.

No common support



Name	Y	D	X
Jake	15	1	6
Gina	10	1	5
Terry	5	1	4
Rosa	10	0	1
Charles	6	0	2
Ray	4	0	3

In this case, none of our control and treated units have common support. Our units are non-comparable in their levels of *X*.

Questions?