Causal Inference

Mixtape Session

MIXTAPE SESSIONS

Roadmap

Randomization inference and causal inference

- “In randomization-based inference, uncertainty in estimates arises naturally from the random assignment of the treatments, rather than from hypothesized sampling from a large population.” (Athey and Imbens 2017)
- Athey and Imbens is part of growing trend of economists using randomization-based methods for doing causal inference
- Unclear (to me) why we are hearing more and more about randomization inference, but we are.
- Could be due to improved computational power and/or the availability of large data instead of samples?

Lady tasting tea experiment

- Ronald Aylmer Fisher (1890-1962)
  - → Two classic books on statistics: *Statistical Methods for Research Workers* (1925) and *The Design of Experiments* (1935), as well as a famous work in genetics, *The Genetical Theory of Natural Science*
  - → Developed many fundamental notions of modern statistics including the theory of randomized experimental design.

Lady tasting tea

- Muriel Bristol (1888-1950)
  - → A PhD scientist back in the days when women weren't PhD scientists
  - → Worked with Fisher at the Rothamsted Experiment Station (which she established) in 1919
  - → During afternoon tea, Muriel claimed she could tell from taste whether the milk was added to the cup before or after the tea
  - → Scientists were incredulous, but Fisher was inspired by her strong claim
  - → He devised a way to test her claim which she passed using randomization inference

Description of the tea-tasting experiment

- Original claim: Given a cup of tea with milk, Bristol claims she can discriminate the order in which the milk and tea were added to the cup
- Experiment: To test her claim, Fisher prepares 8 cups of tea – 4 **milk then tea** and 4 **tea then milk** – and presents each cup to Bristol for a taste test
- Question: How many cups must Bristol correctly identify to convince us of her unusual ability to identify the order in which the milk was poured?
- Fisher’s sharp null: Assume she can’t discriminate. Then what’s the likelihood that random chance was responsible for her answers?

Choosing subsets

- The lady performs the experiment by selecting 4 cups, say, the ones she claims to have had the tea poured first.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

- "8 choose 4" – $\binom{8}{4}$ – ways to choose 4 cups out of 8
  - $\rightarrow$ Numerator is $8 \times 7 \times 6 \times 5 = 1,680$ ways to choose a first cup, a second cup, a third cup, and a fourth cup, in order.
  - $\rightarrow$ Denominator is $4 \times 3 \times 2 \times 1 = 24$ ways to order 4 cups.

Choosing subsets

- There are 70 ways to choose 4 cups out of 8, and therefore a 1.4% probability of producing the correct answer by chance

$$\frac{24}{1680} = 1/70 = 0.014.$$

- For example, the probability that she would correctly identify all 4 cups is $\frac{1}{70}$

Statistical significance

- Suppose the lady correctly identifies all 4 cups. Then ...
  1. Either she has no ability, and has chosen the correct 4 cups purely by chance, or
  2. She has the discriminatory ability she claims.
- Since choosing correctly is highly unlikely in the first case (one chance in 70), the second seems plausible.
- Bristol actually got all four correct
- I wonder if seeing this, any of the scientists present changed their mind

Null hypothesis

- In this example, the null hypothesis is the hypothesis that the lady has no special ability to discriminate between the cups of tea.
- We can never prove the null hypothesis, but the data may provide evidence to reject it.
- In most situations, rejecting the null hypothesis is what we hope to do.

Null hypothesis of no effect

- Randomization inference allows us to make probability calculations revealing whether the treatment assignment was "unusual"
- Fisher's sharp null is when entertain the possibility that no unit has a treatment effect
- This allows us to make "exact" p-values which do not depend on large sample approximations
- It also means the inference is not dependent on any particular distribution (e.g., Gaussian); sometimes called nonparametric

Sidebar: bootstrapping is different

- Sometimes people confuse randomization inference with bootstrapping
- Bootstrapping randomly draws a percent of the total observations for estimation; “uncertainty over the sample”
- Randomization inference randomly reassigns the treatment; “uncertainty over treatment assignment”

(Thanks to Jason Kerwin for helping frame the two against each other)