Announcements

Make sure to sign in on the google form (I send a list of which section questions are useful for which pset questions afterwards)

Pset 11 due Friday 4/28

Causal inference intricacies

Mark whether the following statements are true or false. Explain your answers.

1. $$E(Y|W=0) = E(Y(0)|W=0)$$

2. $$E(Y(0)|W=0) = E(Y(0))$$

3. In an RCT,
$$E(Y(0)|W=0) = E(Y(0))$$

4. In an RCT, $Y \perp\!\!\!\perp W$.

5. $Y = Y(0) + W\tau$ where $\tau = Y(1) - Y(0)$.

6. $$E(Y|W=w) = E(Y(0)) + wE(\tau)$$

7. In an RCT, $$E(Y|W=w) = E(Y(0)) + wE(\tau)$$

8. $E(\tau|W=1) = \frac{E(W\tau)}{E(W)}$

9. Assuming unconfoundedness with another random variable X (conditional on X, $(Y(0), Y(1)) \perp\!\!\!\perp W$), $$E(WY) = E(E(W|X)Y(1))$$

10. Not necessarily assuming unconfoundedness, with $E(\tau|X=x) > 0$ for all x, it is possible to have $E(Y(1)) < E(Y(0))$.

Free Distribution or Cost-Sharing?

This set of questions will be looking at the 2010 paper "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." Before this paper, many development economists argued that cost-sharing, charging a much-reduced but non-zero price for healthcare resources, was necessary to avoid wasting the resources on people who did not need them. In this paper, Jessica Cohen and Pascaline Dupas claim this is not the case. Instead, with a randomized control trial for insecticide treated net (ITN) distribution, they show there is no evidence that cost-sharing reduces wastage, but cost-sharing does significantly decrease demand for ITNs.

The study involved randomizing the cost of ITNs at rural Kenyan health clinics for pregnant women and (1) tracking ITN sales and (2) following up with women to see whether they were using the nets. Originally, four prices were used (\$0, \$0.15, \$0.30, and \$0.60), which represent 100% to 90% subsidies from the original price of the ITN. For simplicity, we will be grouping these into \$0 and non-\$0 groups. The original data is available here.

1. For this first part, for woman i, let W_i be 0 if the woman received a free ITN and 1 if the woman purchased a net. Let Y_i be the indicator of whether the net is hanging when the researchers visit the woman. Suppose we use a finite sample model, so our treatment effects are $\tau_i = y_i(1) - y_i(0)$ and we condition on $y_i(1)$ and $y_i(0)$. The parameter of interest is $\bar{\tau}$, the average treatment effect. Our method of moments estimator is:
$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i W_i}{E(W_i)} - \frac{Y_i(1-W_i)}{E(1-W_i)}$$
where W_i (and therefore Y_i) is the source of randomness. Show that this can be rewritten in terms of n_0 (the number of women who received a free net), n_1 (the number of women who purchased a net), S_0 (the number of women who used a free net), and S_1 (the number of women who used a purchased net).

2. State the Fisher null and explain how to conduct a randomization test. Suppose we use a one-sided test and the usual randomization test procedure. Justify why the reported p-value will never underestimate the true p-value.

3. Because both the W_i and Y_i are binary in this example, we can find the exact distribution of $\hat{\tau}$ under the null. Find this exact distribution.

4. Show how to find the p-value from this distribution.

5. One possible concern whenever implicitly seeking to retain a null is that the desired conclusion can follow from simply not obtaining enough data. Suggest a method to test this.

6. We can follow a similar process to test whether non-zero prices are associated with reduced ITN purchases. Explain why we cannot use the exact test from above. Interpret the results.

```
nsims <- 10^4
sim_store <- vector(length = nsims)
n_free <- sum(sales$cost == 0)
for (i in 1:nsims) {
  # Randomize and compare means
  indices <- sample(1:nrow(sales), n_free, replace = F)
  sim_store[i] <- mean(sales$weeklynetsales[indices]) -
    mean(sales$weeklynetsales[-indices])
}

# Observed value
tau_obs <- mean(sales$weeklynetsales[sales$cost == 0]) -
  mean(sales$weeklynetsales[sales$cost != 0])

# P-value
mean(sim_store >= tau_obs)
```

```
## [1] 0.0391
```

One last thing: When reanalyzing the data to create this set of questions, we made two simplifying assumptions. First, we collapsed the continuous cost variable into a binary. Second, we looked at raw net sales rather than net sales normalized to clinic popularity. Performing this same analysis with normalization actually yields insignificant differences for both the effect of cost on usage and the effect of cost on sales. However, the better way to perform this analysis is to use regression on the continuous cost range, and such analysis yields the conclusions found in the paper whether the net sales are normalized to clinic popularity or not.

Feedback

If you didn't get a chance to do so last week, please take a moment to provide some feedback on section this semester.