Lecture 6

Alex Stephenson

9-8-2021

What did we do last time?

Explained in detail what we mean by randomization and what is required for randomization to eliminate selection bias.

We care about physical procedures that provide us with confidence that treatment is unconfounded with potential outcomes.

What are we doing today?

Announcing that OH this week are immediately after section in the same room

Introducing Randomization Inference

What does inference mean?

It is usually not the case that we are interested in the subject pool exclusively. Rather we are interested in using the units in an experiment to make an inference about a target population.

Tests for inference ought to:

1. Control the false positive rate such that the test false positive size is less than or equal to the test's false positive level. This is satisfied by every common procedure you've heard about.

2. Be unbiased so that the probability of rejecting the null when it is false and the alternative hypothesis is true be at least as great as the probability of rejecting the null hypothesis when it is true and the alternative hypothesis is false.

3. Have the property that as the size of the study population increases while all other factors remain constant $P(Reject|H_0 = FALSE, H_1 = TRUE)$ should go to 1.

Sampling Distributions

A sampling distribution: The frequency distribution of a statistic obtained from hypothetical replications of a randomized experiment.

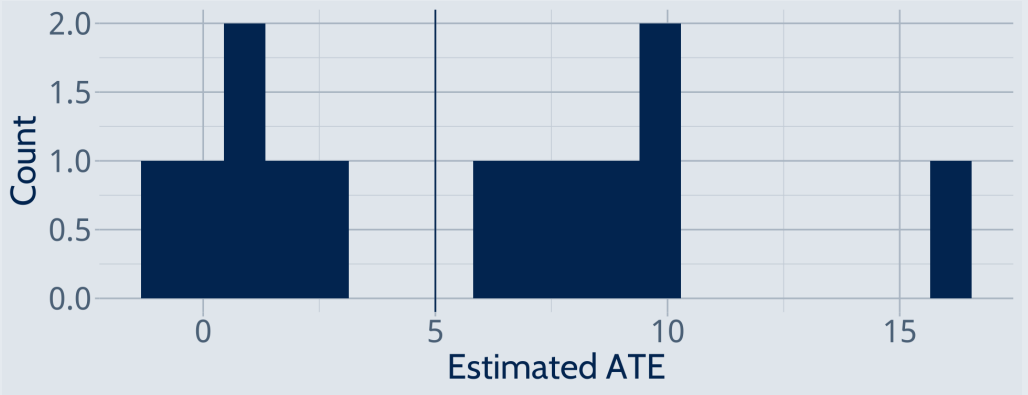
The sampling distribution of the ATE is ~ $N(\mu, \sigma^2)$ as the number of observations in treatment and control groups increase

Example

Suppose we have the following schedule of potential outcomes

```
## # A tibble: 7 × 4
##      id    y0    y1   tau
##   <int> <dbl> <dbl> <dbl>
## 1     1    20    25     5
## 2     2    25    25     0
## 3     3    30    40    10
## 4     4    30    25    -5
## 5     5    20    30    10
## 6     6    25    25     0
## 7     7    25    40    15
```

The sampling distribution of our estimated ATE when assigning 2 units to treatment



In this example, the true ATE is 5 but any single experiment is too high or too low

Definitions

Standard Deviation of a variable X: $\sqrt{\frac{1}{N}\sum_1^N (X_i - \bar{X})^2}$

Sample Standard Deviation plug-in estimator: $\sqrt{\frac{1}{N-1}\sum_1^N (X_i - \bar{X})^2}$

Standard Error: $\sqrt{\frac{1}{J}\sum_1^J (\hat{\theta}_j - \bar{\hat{\theta}})^2}$

Covariance: $\frac{1}{N}\sum_1^N [(Y_i(0) - \bar{Y_i(0)})(Y_i(1) - \bar{Y_i(1)})]$

Properties of Variances and Covariances

The correlation between two variables:

$$\rho[X, Y] = \frac{Cov[X,Y]}{\sigma_x, \sigma_y}$$

Properties of covariances

$$Cov[X, X] = V[X] \geq 0$$

$$Cov[X, X + \tau] = V[X] + Cov(X, \tau)$$

$$Cov[aX, bY] = abCov[X, Y]$$

Proofs of Properties

Some facts about the $E[]$ operator

1. Expectation of a constant $E[c] = c$
2. Expectation of a scalar applied to a variable $E[aX] = aE[X]$
3. Linearity of Expectations $E[aX + bY + c] = aE[X] + bE[Y] + c$

Proofs of Properties

1) $Cov[X, X] = V[X]$

Useful fact 1: to know that the variance can be written as $V[X] = E[X^2] - E[X]^2$

Useful fact 2: The alternative way to write $Cov[X, Y] = E[XY] - E[X]E[Y]$

To prove the first property of interest

$$
\begin{aligned}
Cov[X, X] &= E[XX] - E[X]E[X] \\
Cov[X, X] &= E[X^2] - E[X^2] \\
Cov[X, X] &= V[X]
\end{aligned}
$$

Proofs of Properties

2) $Cov[X, X + \tau] = V[X] + Cov[X, \tau]$

Useful fact: $Cov[X, c] = E[cX] - E[X]E[c] = cE[X] - cE[X] = 0$

Useful fact 2: $V[X, c] = V[X]$ and $V[aX] = a^2V[X]$

To prove the second fact of covariances, provided τ is a constant

$$Cov[X, X + \tau] = Cov[X, X] + Cov[X, \tau]$$
$$Cov[X, X + \tau] = V[X] + Cov[X, \tau]$$
$$Cov[X, X + \tau] = V[X]$$

Proofs of Properties

3) $Cov[aX, bY] = abCov[X, Y]$

To prove this:

$$
\begin{aligned}
Cov[aX, bY] &= E\left[(aX - E[aX])(bY - E[bY])\right] \\
Cov[aX, bY] &= E\left[(aX - aE[X])(bY - bE[Y])\right] \\
Cov[aX, bY] &= E\left[a(X - E[X])b(Y - E[Y])\right] \\
Cov[aX, bY] &= abE[(X - E[X])(Y - E[Y])] \\
Cov[aX, bY] &= abCov[X, Y]
\end{aligned}
$$

Standard Error of the ATE

$$\sigma_{\hat{ATE}} = \sqrt{\frac{1}{N-1}\left(\frac{mV[Y_i(0)])}{N-m} + \frac{(N-m)V[Y_i(1)]}{m} + 2Cov[Y_i(0), Y_i(1)]\right)}$$

Implications for Research Designs

1. As $n \to \infty$, $\sigma_{\hat{ATE}}$ decreases
2. The smaller $V[Y_i(0)]$ or $V[Y_i(1)]$ the smaller the standard error
3. If $Y_i(0)$ and $Y_i(1)$ vary the smaller the covariance between them the smaller the standard error
4. When $V[Y_i(0)]$ and $V[Y_i(1)]$ are similar, assign groups equally. When different, assign more observations to condition with higher variance.
5. Research designs, and experiments in particular, provide information on 4 of the 5 inputs to the formula. We cannot observe the covariance between $Y_i(0)$ and $Y_i(1)$

Conservative Formula for SE in Samples

To get around this, our plug in estimator is conservative and assumes that the treatment effect is constant for all units. This implies that $\rho[Y_i(0), Y_i(1)] = 1$

$$\hat{\sigma}_c = \sqrt{\frac{V[\hat{Y_i}(0)]}{N-m} + \frac{V[\hat{Y_i}(1)]}{m}}$$

where:

$$V[\hat{Y_i}(1)] = \frac{1}{m-1}\sum_{1}^{m}(Y_i(1)|d_i = 1 - \frac{\sum_i^m Y_i(1)|d_i = 1}{m})^2$$

$$V[\hat{Y_i}(0)] = \frac{1}{N-m-1}\sum_{1}^{m}(Y_i(0)|d_i = 0 - \frac{\sum_i^m Y_i(0)|d_i = 0}{N-m})^2$$

Hypothesis Testing with the Sharp Null Hypothesis

Primary question is how likely are we to observe an outcome at least as extreme provided there was no treatment effect at all.

Sharp Null Hypothesis of No Effect: $Y_i(1) = Y_i(0), \forall i$

Null Hypothesis of No Average Effect: The ATE is 0. $\mu_{Y(1)} = \mu_{Y(0)}$

p-values and Randomization Inference

Randomization Inference: The sampling distribution of the test statistic under the null is computed by simulating all (or a large random sample) possible treatment assignment.

p-value: The probability of obtaining a test statistic at least as large as the observed test. For a two tailed test, at least as large means at least as large in absolute value.

Randomization Inference Algorithm

1. Run an experiment
2. Decide the test statistic of interest. $T(D, Y^{obs}, \mathbf{X})$
3. State the null hypothesis of interest
4. State a randomization procedure

Confidence Intervals with Randomization Inference

1. Assume the treatment effect τ_i is the same for all units.
2. Impute missing potential outcomes for each unit.
3. Simulate all possible assignments
4. Sort in ascending order. The CI_α of interest is the $\frac{\alpha}{2}, 1 - \frac{\alpha}{2}$ interval estimates.

Interpretation: Over infinite hypothetical realizations of the experiment, interval have a 95% change of containing the true ATE

Randomization Inference Example

One realization of assigning 2 units to treatment. Why are there only 21 unique ways to do this?

Here we run an experiment and get a plug-in estimate of the ATE of 6.5.

What's the probability that we would obtain an estimate at least as large if the true effect was zero (SN)?

d	y
1	25
0	25
0	30
0	30
0	20
0	25
1	40

Randomization Inference Example

```
## # A tibble: 21 × 1
##         x
##     <dbl>
##  1  -7.5
##  2  -7.5
##  3  -7.5
##  4  -4
##  5  -4
##  6  -4
##  7  -4
##  8  -4
##  9  -0.5
## 10  -0.5
## # … with 11 more rows
```

```
sum(x >= 6.5)/21
```

```
## [1] 0.2380952
```

```
## As a computational trick
# mean(x >=6.5) also works
```

Even if the true value was 0, we'd still expect to see a value of 6.5 or greater ~24% of the time under this allocation scheme.

Do we think it's likely that the true effect is 6.5?

Randomization Inference Example

```
ci_l <- quantile(sort(x), 0.025)
ci_u <- quantile(sort(x), 0.975)

print(paste0("95% CI is (", ci_l,",", ci_u,")"))
```

```
## [1] "95% CI is (-7.5,10)"
```

The confidence interval for this test statistic is very wide and includes 0.

As a rule, if the CI includes 0 we cannot reject the null at the level specified

Next Time

Apply the same methods to block and cluster designs

Compare randomization with approximation methods