QTM 110 Midterm 1 Review

October 3, 2023

1 Types of Questions

- Descriptive: Something we can **see** in the data.
- Predictive: What the **future** might look like
- Causal: If X changes, will Y change?

2 Strength vs. Magnitude

- Covariance: retain units
- Correlation (r): standardized covariance $\Rightarrow r \in [-1, 1]$
  - Sign: direction
  - Absolute Value: strength
- Regression coefficient (β): magnitude $\rightarrow y = \beta x + \alpha$.

All of them should have the same sign: all positive if y increases when x increases; all negative if y decreases when x increases.

- Covariance isn't actually telling us anything because they retain units - they cannot be used in comparison!
- Determine which is the most **predictive**? (***strength**, so we should use correlation coefficient*)
- Determine which is the most effective **chance**? (***magnitude**, so we should consider the regression coefficient*)

3 Counterfactual Dependence

- We need to prove both: (1) when X happens, Y happens; and (2) when X does not happen, Y does not happen.
- Some notations:
  - $T = 0$: control, untreated
  - $T = 1$: treatment, treated
  - Y_1: outcome if treated
  - Y_0: outcome if untreated
- Fundamental problem of casual inference: We cannot observe ATT (Average Treatment Effect of Treated) anyhow!

$$\mathrm{ATT} = \mathbf{E}[Y_1 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 1],$$

  where $\mathbf{E}[Y_0 \mid T = 1]$ is not observable because we cannot have the same group be both treated and untreated.
- Resolution: Estimate the ATT:

$$\widehat{\mathrm{ATT}} = \mathbf{E}[Y_1 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 0],$$

  where both $\mathbf{E}[Y_1 \mid T = 1]$ and $\mathbf{E}[Y_0 \mid T = 0]$ are observable.

- An important equation

$$\boxed{\text{Estimate}} = \boxed{\text{Estimand}} + \text{Bias} + \text{Noise},$$

where possible Estimand's are (remember that Estimand's are unobservable true values in our experiment)

  - $\text{ATE} = \mathbf{E}[Y_1] - \mathbf{E}[Y_0]$
  - $\text{ATT} = \mathbf{E}[Y_1 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 1]$
  - $\text{ATU} = \mathbf{E}[Y_1 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 0]$

The only observable effect (Estimate) is

$$\widehat{\text{ATE}} = \widehat{\text{ATT}} = \widehat{\text{ATU}} = \mathbf{E}[Y_1 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 0]$$

- Following up, we use ATT to represent our general Estimand and $\widehat{\text{ATT}}$ as our Estimate.

4 Identifying Inferential Errors

- Selecting on the DV: Only one **outcome** in data
- Confounder/Common Cause: **omitted variable** (We are not comparing apples to apples)

5 Bias and Noise

- Bias: Systematic error; Not apples-to-apples; Inaccuracy

$$\text{Bias} = \mathbf{E}[Y_0 \mid T = 1] - \mathbf{E}[Y_0 \mid T = 0]$$

  - To measure Bias, we give the two groups Placebo (however, bias is still unobservable).
  - Eliminating bias: **randomization**.

- Noise: Random error; $\mathbf{E}[\text{Noise} = 0]$; Imprecision
  - If we repeat our experiment for enough times, our Noise should be 0.
  - We are not super stressful with Noise since the solution is to increase sample size.

6 Confidence Interval and p-values

- Point Estimate (=$\widehat{\text{ATT}}$) $\xrightarrow[\pm 2\times\text{SE}]{\text{MATH}}$ 95% CI
  - If we run the experiment identically 100 times, 95 of them should contain the true value of the estimand.
  - When we have 99% CI, then we should find larger intervals.

- Interpreting a CI: Sign and Magnitude (every important if the interval contains 0).
  - $\widehat{\text{ATT}} = 10;\ 95\%\ \text{CI} = [6, 14]$: If we run the experiment identically 100 times and calculation the confidence interval in each repetition, out of 95 times, we are confident that $[6, 14]$ captures the ATT. Furthermore, ATT is likely **positive** and has a **large magnitude**.
  - $\widehat{\text{ATT}} = 3;\ 95\%\ \text{CI} = [-2, 8]$: If we run the experiment identically 100 times, and we calculate the confidence interval in each repetition, out of 95 times, we are confident that the interval $[-2, 8]$ captures the ATT. However, ATT **sign is unclear** and has a **small magnitude**.

- Null Hypothesis: $\text{ATT} = 0$. ($\Rightarrow$ Any nonzero $\widehat{\text{ATT}}$ is due to noise)
  - **Fail to reject the null**: $\widehat{\text{ATT}}$ could be due to noise
  - **Reject the null**: $\widehat{\text{ATT}}$ is probably not due to noise

- p-values:

- High p-value $\rightarrow$ **High** chance of observing if $\text{ATT} = 0 \rightarrow$ **fail to reject** the null $\rightarrow$ ATT could **reasonably** be 0.
- Low p-value $\rightarrow$ **Low** chance of observing if $\text{ATT} = 0 \rightarrow$ **reject** the null $\rightarrow$ ATT is **unlikely** be 0.

- Threshold of p-values: 0.05
- Connecting p-values and CI's
  - $\widehat{\text{ATT}} = 5;\ 95\%\ \text{CI} = [-1, 11];\ p$-value $= 0.13$: **fail to reject** the null, ATT could **reasonably** be 0 (the result could be due to noise).
  - $\widehat{\text{ATT}} = 5;\ 95\%\ \text{CI} = [2, 8];\ p$-value $= 0.03$: **reject** the null, ATT is **unlikely** be 0 (the result is not due to noise).