PMAP 8131 Applied Research Methods

Sampling

Matteo Zullo

Georgia State University & Georgia Institute of Technology

June 27, 2022

Outline

1. Hypothesis Testing
   - Central Limit Theorem
   - A/B testing
2. Randomized Controlled Trials
   - Potential Outcomes
   - Confounding
3. Power Analysis
   - Effect sizes
   - Clustered Randomized Trials

Table of Contents

Hypothesis testing

Central limit theorem

- Property of the **mean** of a distribution, not of distribution
- N is number of **samples**, not population size
- Applies to **any distribution X** (normal, exponential, etc.)

$$\lim_{N \to \infty} \overline{X} \sim (\mu_X, \sqrt{\frac{\sigma_X^2}{N}})$$

Hypothesis testing

Central limit theorem: Simulation

- Lognormal with mode = 8 and median = 10
- One population of size 1,000

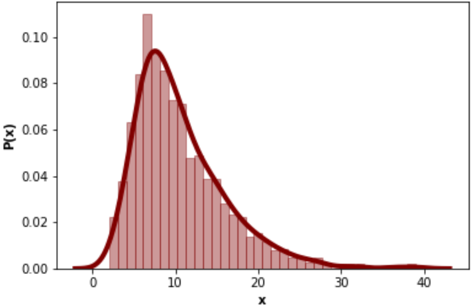

Figure: Lognormal distribution of size 1,000

Hypothesis testing

Central limit theorem: Simulation

- Lognormal with mode = 8 and median = 10
- 10,000 populations of size 1,000 ($1,000 \times 10,000$)

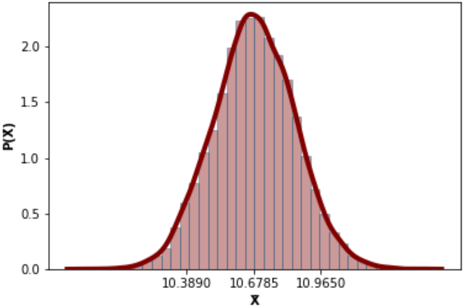

Figure: Distribution of means of 10,000 lognormals with 95% CI

Hypothesis testing

Example: Churn

- A clothing brand has revamped its fashion line in the hope to **reduce consumer churn**. Churn dropped down from 40% to 30% in a batch of 100 clients purchasing from the new line.
  Is that enough observations to green light the new line?

Hypothesis testing

Example: Churn

- Null Hypothesis (H_0):
- Alternative Hypothesis (H_1):
- Significance level (α): 5%
- Type of test:

Hypothesis testing

Example: Churn

- Null Hypothesis (H_0): Callback rate 40%
- Alternative Hypothesis (H_1): Callback rate less $< 40\%$
- Significance level (α): 5%
- Type of test: One-tailed

Hypothesis testing

Example: Churn

- Mean and variance of the callback distribution

$$\begin{aligned}\hat{\mu} =& p = 0.3 \\ \hat{\sigma}^2 =& p(1-p) = 0.3(0.7) = 0.21\end{aligned}$$

- Confidence interval for population parameter

,

Hypothesis testing

Example: Churn

- Mean and variance of the callback distribution

$$\begin{aligned}\hat{\mu} =& p = 0.3\\ \hat{\sigma}^2 =& p(1-p) = 0.3(0.7) = 0.21\end{aligned}$$

- Confidence interval for population parameter

$$\begin{aligned}CI =& 0.3 \pm t_{\alpha}\sqrt{\frac{\hat{\sigma}^2}{N}}\\ =& 0.3 \pm 1.645\sqrt{\frac{0.21}{100}} = [0.22, 0.38]\end{aligned}$$

,

Hypothesis testing

Example: Churn

- Mean and variance of the callback distribution

$$\begin{aligned}\hat{\mu} =& p = 0.3\\ \hat{\sigma}^2 =& p(1-p) = 0.3(0.7) = 0.21\end{aligned}$$

- Confidence interval for population parameter

$$\begin{aligned}CI =& 0.3 \pm t_\alpha \sqrt{\frac{\hat{\sigma}^2}{N}}\\ =& 0.3 \pm 1.645 \sqrt{\frac{0.21}{100}} = [0.22, 0.38]\end{aligned}$$

- The CI does not cover 0.40, so <u>100 customers is enough</u>

,

Table of Contents

Randomized Controlled Trials

Randomization: The "gold standard" for evaluation

- Average Treatment on the Treated (ATT)

$$ATT = \underbrace{E(Y^1|T=1,X,\lambda)}_{\text{outcomes of treated, as if treated}} - \underbrace{E(Y^0|T=1,X,\lambda)}_{\text{outcomes of treated, as if not treated}}$$

- Average Treatment Effect (ATE)

$$ATE = \underbrace{E(Y^1|T=1,X,\lambda)}_{\text{outcomes of treated, as if treated}} - \underbrace{E(Y^0|T=0,X,\lambda)}_{\text{outcomes of controls, as if not treated}}$$

Randomized Controlled Trials

Randomization: The "gold standard" for evaluation

- Average Treatment on the Treated (ATT)

$$ATT = \underbrace{E(Y^1|T=1,X,\lambda)}_{\text{outcomes of treated, as if treated}} - \underbrace{E(Y^0|T=1,X,\lambda)}_{\text{outcomes of treated, as if not treated}}$$

- Average Treatment Effect (ATE)

$$ATE = \underbrace{E(Y^1|T=1,X,\lambda)}_{\text{outcomes of treated, as if treated}} - \underbrace{E(Y^0|T=0,X,\lambda)}_{\text{outcomes of controls, as if not treated}}$$

- Perfect randomization
  - Removes bias from observables and unobservables

Randomized Controlled Trials

Confounding: Facebook example

ID	Married	FB account
001	1	1
002	1	0
003	1	1
004	1	1
005	0	0
006	0	0
007	1	1
008	0	0

Table: Effect of Facebook usage on marriages

Randomized Controlled Trials

Confounding: Facebook example

ID	Married	FB account	Extraversion
001	1	1	High
002	1	0	Medium
003	1	1	High
004	1	1	High
005	0	0	Low
006	0	0	Low
007	1	1	Medium
008	0	0	Low

Table: Effect of Facebook usage on marriages

Randomized Controlled Trials

Confounding: Facebook example

- Biased estimate

$$ATT = P(Marry|FB = 1) - P(Marry|FB = 0)$$

- Unbiased estimate

$$ATT = P(Marry|FB = 1, \textbf{Extra}) - P(Marry|FB = 0, \textbf{Extra})$$

Randomized Controlled Trials

- RCT: Treated and controls the same via randomization

$$E(Y^0|T = 0, X, \lambda) = E(Y^0|T = 1, X, \lambda)$$
$$E(Y^1|T = 1, X, \lambda) = E(Y^1|T = 0, X, \lambda)$$

- It follows that $ATT = ATE$

Randomized Controlled Trials

Confounding: Definition

- A confounder affects outcome and treatment alike

$$\rho(T, X) \neq 0$$
$$\rho(Y, X) \neq 0$$

- A confounder, when omitted, biases ATT and ATE
  - If $\rho(T, X) > 0$ and $\rho(Y, X) > 0$, upward bias
  - If $\rho(T, X) > 0$ and $\rho(Y, X) < 0$, downward bias

Randomized Controlled Trials

Confounders: *There's one for every season*

- Observable
  - Parental income (Economics)
  - Marital status (Sociology)
- Unobservable
  - Ability (Education)
  - Ideology (Political Science)

Randomized Controlled Trials

RCT checklist

1. Preparing trial
   - **Debiasing ✓**
   - **Sampling ✓**
2. Conducting trial
   - Attrition
   - Interaction

Table of Contents

Power Analysis

Effect sizes: Definition

- A **standardized measure of association**

Power calculations

- Bivariate associations: Closed-form solutions
- Multivariate associations: Simulation-based solutions

Power Analysis

- Cohen's d:

$$d = \frac{\mu_T - \mu_C}{\sigma}$$

- with pooled variance $\sigma = \sqrt{\frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T + n_C - 1}}$:
  - n_T and n_C = number of treatments and controls
  - s_T^2 and s_C^2 = variances of treatments and controls

Power Analysis

- Effect sizes: Rule of thumb
  - < 0.2: *very small*
  - $0.2 - 0.5$: *small*
  - $0.5 - 0.8$: *large*
  - > 0.8: *very large*

Power Analysis

Bivariate associations

- Minimum Detectable Effect Size (MDE)

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\beta})\sqrt{\frac{\sigma^2}{n_T n_C / N}}$$

- where:
    - α = probability of type I error
    - β = probability of type II error
    - σ = variance of the outcome (assume $\sigma_T^2 = \sigma_C^2$)
    - N = total number of units ($N = n_T + n_C$)
- *MDE* decreases when α, β, and σ increase, and decreases when N increases

Power Analysis

Bivariate associations

1. Choose α and β
   - Standard values: $\alpha = 0.05$, $\beta = 0.8$ (Cohen, 1988)
2. Choose n_1 and n_2
   - Optimal treatment ratio: $\frac{n_T}{N} = \frac{1}{1+\frac{\sigma_{control}}{\sigma_{treat}}}$
   - With budget contraint: $\frac{n_T}{n_C} = \sqrt{\frac{c_C}{c_T}}$ (Duflo, 2007)
3. Set σ following prior literature or similar population
4. Calculate MDE or solve for N choosing a target MDE

Power Analysis

Multivariate associations: Variance of OLS estimator

$$Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{S_{XX}(1 - R_j^2)}$$

- where:
  - $\hat{\sigma}^2 =$ residual error
  - $S_{XX} =$ variation in the predictor
  - $R_j^2 = R^2$ from regression of X_j on other predictors
- Multicollinearity
  - As R_j^2 increases, variance of OLS estimator increases

Power Analysis

- Adding regressors might increase variance
  - If residual variance SSE (numerator) less than new degrees of freedom p in the model (denominator)

$$\sigma^2 = \frac{SSE}{n-p}$$

  - If residual error $\hat{\sigma}^2$ (numerator) reduces less than new multicollinearity R_j^2 with other regressors (denominator)

$$Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j(1-R_j^2)}$$

Power Analysis

Stratification

- Example: Categorical regressor

ID	SAT	Treat
001	1300	0
002	1550	0
003	1400	1
004	1500	1

$\Longrightarrow$

ID	SAT	Treat	Female
001	1300	0	0
002	1550	0	1
003	1400	1	0
004	1500	1	1

- Adding *Female* increases dimensionality from 2 to 4
  - Before: treat, control
  - After: M treat, M control, F treat, F control

Power Analysis

Stratification

- Adding regressors increases dimensionality
  - Categorical: Add $df \times k$ dimensions (k levels)
  - Continuous: Add infinite dimensions
- Idea
  - Reproduce optimal treatment ratio within each subgroup

Power Analysis

Clustered Randomized Trials: School example

student_id	prep_class	SAT
001	0	1,300
002	1	1,500
003	0	1,200
004	1	1,400
005	0	900
006	0	1,000

Power Analysis

Clustered Randomized Trials: School example

student_id	school_id	prep_class	SAT
001	A	0	1,300
002	A	1	1,500
003	A	0	1,200
004	B	1	1,400
005	B	0	900
006	B	0	1,000

Power Analysis

Intra-class correlation coefficient (ICC)

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$$

- where:
  - $\sigma^2_{between}$ = variance across clusters
  - σ^2_{within} = variance within clusters

Power Analysis

- For K clusters, M subjects in each cluster
  - Effective sample size (ESS)

$$\overline{N} = \frac{M \cdot K}{DE}$$

  - Design Effect (DE)

$$DE = 1 + (M-1)ICC$$

  - Minimum Detectable Effect Size (MDE)

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\beta})\sqrt{\frac{\sigma^2}{n_T n_C / \overline{N}}}$$

Power Analysis

Clustered Randomized Trials: School example

- Without clustering (i.e., ICC = 0)

$$N = \frac{2 \cdot 3}{1} = 6$$

- With clustering and perfect correlation (i.e., ICC = 1)

$$\overline{N} = \frac{2 \cdot 3}{1 + ICC(3-1)} = \frac{6}{1+2} = 2$$

- Hence:

$$K \leq ESS \leq N$$

Power Analysis

- When clusters highly correlated ($\sigma^2_{between} >> 0$), $ESS \downarrow$
  - Recruit more subjects
  - Increase cluster size or number of clusters
  - Use simulation software (G*Power)
- When clusters uncorrelated ($\sigma^2_{between} \approx 0$), $ESS \approx N$
- When clusters are unbalanced, not much of a deal