Inference for randomized experiments.

The Average Treatment Effect (ATE). [also Average Causal Effect (ACE)]

$$\tau = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i(1) - Y_i(0)\right).$$

This is a fixed quantity. b/c the potential outcomes are fixed.

This is unidentifiable, can only estimate it.

Estimator for τ: The Neyman estimator / difference-in-means.

$$\hat{\tau} := \frac{1}{n_1}\sum_{Z_i=1} Y_{i,obs} - \frac{1}{n_0}\sum_{Z_i=0} Y_{i,obs}$$

Rem: $\hat{\tau} = \frac{1}{n_1}\sum_{Z_i=1} Y_i(1) - \frac{1}{n_0}\sum_{Z_i=0} Y_i(0).$

Properties:

① $\mathbb{E}[\hat{\tau}] = \tau.$

② $\mathrm{Var}(\hat{\tau}) \leq \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0}$, where,

for $k=0,1$, $S_k^2 = \cancel{\frac{1}{n_k-1}\sum_{i=1}^{n_k}\left(Y_i(k) - \bar{Y}(k)\right)^2}.$

$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i(k) - \bar{Y}(k)\right)^2$, $\bar{Y}(k) = \frac{1}{n}\sum_{i=1}^{n} Y_i(k)$

Pf of ①:

$$\hat{\tau} = \frac{1}{n_1}\sum_{Z_i=1} Y_i(1) - \frac{1}{n_0}\sum_{Z_i=0} Y_i(0).$$

$$= \frac{1}{n_1}\sum_{i=1}^{n} Z_i \cdot Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)\cdot Y_i(0).$$

$$\mathbb{E}[\hat{\tau}] = \sum_{i=1}^{n} \mathbb{E}\left[\frac{Z_i}{n_1} \cdot Y_i(1)\right] - \sum_{i=1}^{n} \mathbb{E}\left[\frac{1-Z_i}{n_0} \cdot Y_i(0)\right].$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\frac{Z_i}{n_1}\right] \cdot Y_i(1) - \sum_{i=1}^{n} \mathbb{E}\left[\frac{1-Z_i}{n_0}\right] \cdot Y_i(0).$$

$$\mathbb{P}(Z_i = 1) = \frac{n_1}{n}, \quad \mathbb{P}(Z_i = 0) = \frac{n_0}{n}.$$

$$= \sum_{i=1}^{n} \frac{n_1/n}{n_1} Y_i(1) - \sum_{i=1}^{n} \frac{1}{n} \cdot Y_i(0).$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(Y_i(1) - Y_i(0)\right).$$

$$= \tau. \qquad \blacksquare$$

Confidence intervals

$$V_{\text{Neyman}} = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0}$$

Want a sample estimate of V_{Neyman}.

Define the estimator $\hat{V}_{\text{Neyman}} = \frac{s_1^2}{n_1} + \frac{s_0^2}{n_0}$,

where $s_1^2 = \frac{1}{n_1 - 1} \cdot \sum_{Z_i=1} (Y_{i,obs} - \bar{Y}_{obs,t})^2$, $\quad \bar{Y}_{obs,t} = \frac{1}{n_1 - 1} \sum_{Z_i=1} Y_{i,obs}$

$s_0^2 = \ldots$

[s_k^2 = sample estimates of S_k^2].

lowercase s_k — upper case S_k defined in (*)

Fact. $\mathbb{E}[\hat{V}_{\text{Neyman}}] = V_{\text{Neyman}}$.

Pf. Ex.....

Thm. Under regularity conditions, $\frac{\hat{\tau} - \tau}{\sqrt{\hat{V}_{\text{Neyman}}}} \Rightarrow N(0, \sigma^2)$,

where $\sigma^2 \leq 1$

Hence, an asymptotically valid 95% CI for the ATE τ

is given by $(\hat{\tau} - 1.96\sqrt{\hat{V}_{Neyman}},\ \hat{\tau} + 1.96\sqrt{V_{Neyman}})$.

Hypothesis testing

Neyman's weak null H_{ON} : $ATE = 0$.

Fisher's strong null H_{OF} : $ITE_i = Y_i(1) - Y_i(0) = 0 \ \forall i$.

To test H_{ON}, consider the test statistic

$$t = \frac{\hat{\tau}}{\sqrt{\hat{V}_N}}$$

Get asymptotically valid, conservative P-value by comparing t against $N(0,1)$.

To test H_{OF}, observe that if H_{OF} is true, we know all the entries in the Science Table. Hence, we can use permutation test.

Fisher's exact test.

Assume binary outcome variable Y.
E.g. Z = take aspirin.
Y = fever goes away.

Science Table:

$Y_i(0)$	$Y_i(1)$
1	
	1
	0

	0
1	
0	
1	
⋮	

Can summarize by cross tabulation, to get a 2×2 contingency table.

	C	T	
F	n_{00}	n_{01}	n_F
S	n_{10}	n_{11}	n_S.
	n_0	n_1	

Assume H_{0F} holds. Then we know all entries in Science Table,

$Y_i(0)$	$Y_i(1)$
1	1
1	1
0	0
0	0
1	1
0	0
1	1
⋮	

n_S = no. of rows with 1.
n_F = no. of rows with 0.

Also, n_{00}, n_{01}, n_{10}, n_{11} all follow hypergeometric distributions.

Use this to compute p-value for observed

$n_{00}, n_{01}, n_{10}, n_{11}$.