Introduction to Causal Inference

Lecture 3

Fisher's Exact P-Values & Neyman's Repeated Sampling Approach for Completely Randomized Experiments

邓婉璐

WANLUDENG@TSINGHUA.EDU.CN



清华大学统计与数据科学系

The Honey Experiment Data (2007)

Unit: $N = 72$ children
Treatment: a single dose of buckwheat honey
Control: no active treatment
Two outcomes:
cough frequency afterwards (cfa) & cough severity afterwards (csa)
Two covariates:
cough frequency prior (cfp) & cough severity prior (csp)
Completely randomized assignment:
$N_t = 35$ & $N_c = 37$

Table 5.1. ***Summary Statistics for Observed Honey Data***

Variable	Mean	(S.D.)	Mean Controls	Mean Treated
Cough frequency prior to treatment (cfp)	3.86	(0.92)	3.73	4.00
Cough frequency after treatment (cfa)	2.47	(1.61)	2.81	2.11
Cough severity prior to treatment (csp)	3.99	(1.03)	3.97	4.00
Cough severity after treatment (csa)	2.54	(1.74)	2.86	2.20



Outcomes are measured on a scale from 0 ("not at all frequent/severe") to 6 ("extremely frequent/severe").

Completely Randomized Experiment

- A leading case of a classical randomized experiment
- The number of treated units N_t is fixed
- N_t units are randomly selected, from a population of N units, to receive the active treatment, with the remaining N_c assigned to the control group.
- Each unit has unit assignment probability $q = N_t/N$
- The assignment mechanism:

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 1 \Big/ \binom{N}{N_t} & \text{if } \sum_{i=1}^{N} W_i = N_t, \\ 0 & \text{otherwise,} \end{cases}$$

$$\binom{N}{N_t} = \frac{N!}{N_t!(N-N_t)!}$$

- subset of the set of possible values for $\mathbf{W}$ with positive probability

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i=1}^{N} W_i = N_t \right\}$$



Outline

- Fisher's Exact P-Values
  - *whether there exists a non-zero unit-level causal effect?*
- Neyman's Repeated Sampling Approach
  - *how to estimate the average causal effect?*
  - Finite Sample
  - Super Population
- Summary
- Appendix



Fisher’s approach

WHETHER THERE EXISTS A NON-ZERO UNIT-LEVEL CAUSAL EFFECT?



A Simple Example with 6 Units

Table 5.3. ***Cough Frequency for the First Six Units from the Honey Study***

Unit	Potential Outcomes				
	Cough Frequency (cfa)		Observed Variables		
	$Y_i(0)$	$Y_i(1)$	W_i	X_i (cfp)	Y_i^{obs} (cfa)
1	?	3	1	4	3
2	?	5	1	6	5
3	?	0	1	4	0
4	4	?	0	4	4
5	0	?	0	1	0
6	1	?	0	5	1

The key challenge of causal inference: half of the potential outcomes are missing!



Idea(1): Fisher's Sharp Null Hypothesis

$$H_0: \tau_{fs} = 0$$

?

$$H_0: \quad Y_i(0) = Y_i(1) \quad \text{for} \quad i = 1, \ldots, 6.$$

- no effect of the active versus control treatment for each unit
- we can fill in all the missing potential outcomes

Idea(1): Fill in Missing Potential Outcomes based on FSNH

Table 5.3. ***Cough Frequency for the First Six Units from the Honey Study***

Unit	Potential Outcomes				
	Cough Frequency (cfa)		Observed Variables		
	$Y_i(0)$	$Y_i(1)$	W_i	X_i (cfp)	Y_i^{obs} (cfa)
1	?	3	1	4	3
2	?	5	1	6	5
3	?	0	1	4	0
4	4	?	0	4	4
5	0	?	0	1	0
6	1	?	0	5	1

$$H_0: \quad Y_i(0) = Y_i(1) \quad \text{for} \quad i = 1, \ldots, 6.$$

Unit	Potential Outcomes					
	Cough Frequency (cfa)		Observed Variables			
	$Y_i(0)$	$Y_i(1)$	Treatment	X_i	Y_i^{obs}	rank(Y_i^{obs})
1	(3)	3	1	4	3	4
2	(5)	5	1	6	5	6
3	(0)	0	1	4	0	1.5
4	4	(4)	0	4	4	5
5	0	(0)	0	1	0	1.5
6	1	(1)	0	5	1	3

Idea(2): Find a Test Statistic

A function of assignment **W** under Fisher's sharp null hypothesis

$$T(\mathbf{W}, \mathbf{Y}^{\text{obs}}) = T^{\text{dif}} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}$$

$$\overline{Y}_{\text{t}}^{\text{obs}} = \sum_{i:W_i=1} Y_i^{\text{obs}}/N_{\text{t}} \qquad \overline{Y}_{\text{c}}^{\text{obs}} = \sum_{i:W_i=0} Y_i^{\text{obs}}/N_{\text{c}}$$

$$N_{\text{t}} = \sum_{i=1}^{N} W_i \qquad N_{\text{c}} = \sum_{i=1}^{N} (1 - W_i)$$

Idea(2): Find a Test Statistic

Unit	Potential Outcomes				
	Cough Frequency (cfa)		Observed Variables		
	$Y_i(0)$	$Y_i(1)$	Treatment	X_i	Y_i^{obs}
1			1	4	3
2			1	6	5
3			1	4	0
4			0	4	4
5			0	1	0
6			0	5	1

$$T^{\text{obs}} = T(\mathbf{W}, \mathbf{Y}^{\text{obs}}) = \bar{Y}_t^{obs} - \bar{Y}_c^{obs} \quad \Leftarrow \quad \mathbf{W}^{\text{obs}} = (1,1,1,0,0,0)$$

$$= |(Y_1^{\text{obs}} + Y_2^{\text{obs}} + Y_3^{\text{obs}})/3 - (Y_4^{\text{obs}} + Y_5^{\text{obs}} + Y_6^{\text{obs}})/3| = |8/3 - 5/3| = 1.00.$$

$$T(\tilde{\mathbf{W}}, \mathbf{Y}^{\text{obs}}) = (Y_2^{\text{obs}} + Y_3^{\text{obs}} + Y_6^{obs})/3 - (Y_1^{\text{obs}} + Y_4^{\text{obs}} + Y_5^{obs})/3$$

$$= 6/3 - 7/3 = -0.33 \quad \Leftarrow \quad \tilde{\mathbf{W}} = (0,1,1,0,0,1)$$

Idea(3): Enumerate all Possible Values of W

N=6 does NOT have enough power to reject H_0

- P-value:16/20 = 0. 80
- In 16 rows, $T^{dif} \geq 1$

- Key question: How unusual or extreme is the observed test statistics (i.e., 1) assuming the null hypothesis is true?

- The null distribution of test statistics T^{dif} under Fisher's sharp null hypothesis
- The randomness is introduced by the assignment vector W

There are C_6^3 =20 different possible assignment vectors.

Table 5.5. ***Randomization Distribution for Two Statistics for the Honey Data from Table 5.3***

W_1	W_2	W_3	W_4	W_5	W_6	Statistic: Absolute Value of Difference in Average $\bar{Y}_t^{obs} - \bar{Y}_c^{obs}$
0	0	0	1	1	1	−1.00
0	0	1	0	1	1	−3.67
0	0	1	1	0	1	−1.00
0	0	1	1	1	0	−1.67
0	1	0	0	1	1	−0.33
0	1	0	1	0	1	2.33
0	1	0	1	1	0	1.67
0	1	1	0	0	1	−0.33
0	1	1	0	1	0	−1.00
0	1	1	1	0	0	1.67
1	0	0	0	1	1	−1.67
1	0	0	1	0	1	1.00
1	0	0	1	1	0	0.33
1	0	1	0	0	1	−1.67
1	0	1	0	1	0	−2.33
1	0	1	1	0	0	0.33
1	1	0	0	0	1	1.67
1	1	0	0	1	0	1.00
1	1	0	1	0	0	3.67
1	1	1	0	0	0	**1.00**

Three Steps to Get Fisher's Exact P-value

1. specifying the sharp null hypothesis
   - $H_0 : Y_i(1) = Y_i(0) + C_i \quad \text{for } i = 1, \ldots, N.$
   - $H_0 : Y_i(1)/Y_i(0) = C_i \quad \text{for } i = 1, \ldots, N.$

   $H_0 : Y_i(1) = Y_i(0) + C_i \quad \text{for } i = 1, \ldots, N.$

   For constant additive /multiplicative treatment effect
2. choosing the test statistic T

   All data we can observe in causal inference
   - a known, real-valued function $T(\boldsymbol{W}, \boldsymbol{Y}^{obs}, \boldsymbol{X})$
3. measuring the extremeness / Computation of P-Values
   - based on the exact null distribution of T under H_0 obtained by exhaustive enumeration

Step 2. Choosing the Test Statistic T (1)

$H_0: Y_i(1) = Y_i(0)$

$$T^{\text{dif}} = \left|\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right| = \left|\frac{\sum_{i:W_i=1} Y_i^{\text{obs}}}{N_{\text{t}}} - \frac{\sum_{i:W_i=0} Y_i^{\text{obs}}}{N_{\text{c}}}\right|$$

$H_1: Y_i(1) = Y_i(0) + C$

$$T^{\text{log}} = \left|\frac{\sum_{i:W_i=1} \ln(Y_i^{\text{obs}})}{N_{\text{t}}} - \frac{\sum_{i:W_i=0} \ln(Y_i^{\text{obs}})}{N_{\text{c}}}\right|$$

$H_1: Y_i(1) = Y_i(0) \cdot C$

$$T^{\text{median}} = \left|\text{med}_t(Y_i^{\text{obs}}) - \text{med}_c(Y_i^{\text{obs}})\right| \qquad T^{\text{quant}} = \left|q_{\delta,t}(Y_i^{\text{obs}}) - q_{\delta,c}(Y_i^{\text{obs}})\right|$$

insensitive to outliers

$$T^{\text{t-stat}} = \left|\frac{\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}}{\sqrt{s_c^2/N_{\text{c}} + s_t^2/N_{\text{t}}}}\right|$$

$$s_t^2 = \sum_{i:W_i=1} (Y_i^{\text{obs}} - \overline{Y}_{\text{t}}^{\text{obs}})^2/(N_{\text{t}} - 1)$$

$$s_c^2 = \sum_{i:W_i=0} (Y_i^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}})^2/(N_{\text{c}} - 1)$$

unequal variances in the two groups

$$T^{\text{rank}} = \left|\overline{R}_t - \overline{R}_c\right| = \left|\frac{\sum_{i:W_i=1} R_i}{N_{\text{t}}} - \frac{\sum_{i:W_i=0} R_i}{N_{\text{c}}}\right|$$

insensitive to outliers or skewed distributions

$$R_i = R_i(Y_1^{\text{obs}}, \ldots, Y_N^{\text{obs}}) = \sum_{j=1}^{N} \mathbf{1}_{Y_j^{\text{obs}} < Y_i^{\text{obs}}} + \frac{1}{2}\left(1 + \sum_{j=1}^{N} \mathbf{1}_{Y_j^{\text{obs}} = Y_i^{\text{obs}}}\right) - \frac{N+1}{2}.$$

Step 2. Choosing T (1): Rank Test Statistic

Unit	Potential Outcomes					
	Cough Frequency (cfa)		Observed Variables			
	$Y_i(0)$	$Y_i(1)$	Treatment	X_i	Y_i^{obs}	rank(Y_i^{obs})
1	(3)	3	1	4	3	4
2	(5)	5	1	6	5	6
3	(0)	0	1	4	0	1.5
4	4	(4)	0	4	4	5
5	0	(0)	0	1	0	1.5
6	1	(1)	0	5	1	3

$$T^{\text{rank}} = \left|\overline{R}_t - \overline{R}_c\right| = \left|\frac{\sum_{i:W_i=1} R_i}{N_t} - \frac{\sum_{i:W_i=0} R_i}{N_c}\right| \leftarrow \mathbf{W}^{\text{obs}} = (1,1,1,0,0,0)$$

Step 2. Choosing T (1): Rank Test Statistic

- P-value:16/20 = 0. 80
- In 16 rows, $T^{rank} \geq 0.67$

- How unusual or extreme is the observed test statistics assuming the null hypothesis is true?

- The null distribution of test statistics T^{rank} under Fisher's sharp null hypothesis
- The randomness is introduced by the assignment vector W

There are C_6^3 =20 different possible assignment vectors.

Table 5.5. ***Randomization Distribution for Two Statistics for the Honey Data from Table 5.3***

						Statistic: Absolute Value of Difference in Average	
W_1	W_2	W_3	W_4	W_5	W_6	Levels (Y_i)	Ranks (R_i)
0	0	0	1	1	1	−1.00	−0.67
0	0	1	0	1	1	−3.67	−3.00
0	0	1	1	0	1	−1.00	−0.67
0	0	1	1	1	0	−1.67	−1.67
0	1	0	0	1	1	−0.33	0.00
0	1	0	1	0	1	2.33	2.33
0	1	0	1	1	0	1.67	1.33
0	1	1	0	0	1	−0.33	0.00
0	1	1	0	1	0	−1.00	−1.00
0	1	1	1	0	0	1.67	1.33
1	0	0	0	1	1	−1.67	−1.33
1	0	0	1	0	1	1.00	1.00
1	0	0	1	1	0	0.33	0.00
1	0	1	0	0	1	−1.67	−1.33
1	0	1	0	1	0	−2.33	−2.33
1	0	1	1	0	0	0.33	0.00
1	1	0	0	0	1	1.67	1.67
1	1	0	0	1	0	1.00	0.67
1	1	0	1	0	0	3.67	3.00
1	1	1	0	0	0	**1.00**	**0.67**

Step 2. Choosing T (2)

- Model-Based Statistics (parametric model)
  - $Y_i(0) \sim N(\mu_c, \sigma_c^2),\ Y_i(1) \sim N(\mu_t, \sigma_t^2)$

$$T^{\text{model}} = \left|\hat{\mu}_t - \hat{\mu}_c\right| = \left|\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right| = T^{\text{dif}}$$

  - $\log Y_i(0) \sim N(\mu_c, \sigma_c^2), \log Y_i(1) \sim N(\mu_t, \sigma_t^2)$

$$T^{\text{model}} = \left|\hat{\mu}_{\text{ml},t} - \hat{\mu}_{\text{ml},c}\right|$$

MLE of the mean parameter

good approximations to distribution of the potential outcomes under some alternative hypothesis

- The Kolmogorov-Smirnov Statistic (non-parametric model)

$$T^{\text{ks}} = \sup_y \left|\hat{F}_t(y) - \hat{F}_c(y)\right| = \max_{i=1,\ldots,N} \left|\hat{F}_t\left(Y_i^{\text{obs}}\right) - \hat{F}_c\left(Y_i^{\text{obs}}\right)\right|$$

$$\hat{F}_c(y) = \frac{1}{N_{\text{c}}} \sum_{i:W_i=0} \mathbf{1}_{Y_i^{\text{obs}} \le y}, \quad \text{and} \quad \hat{F}_t(y) = \frac{1}{N_{\text{t}}} \sum_{i:W_i=1} \mathbf{1}_{Y_i^{\text{obs}} \le y},$$

Detect any differences in distributions

Step 2. Choosing T: A Simulation Study

$T^{\text{dif}}(\text{-}), T^{\text{median}}(\text{—}), T^{\text{rank}}(\ldots)$

$\alpha = 0.1$

Figure 5.1. Additive model with normal outcomes

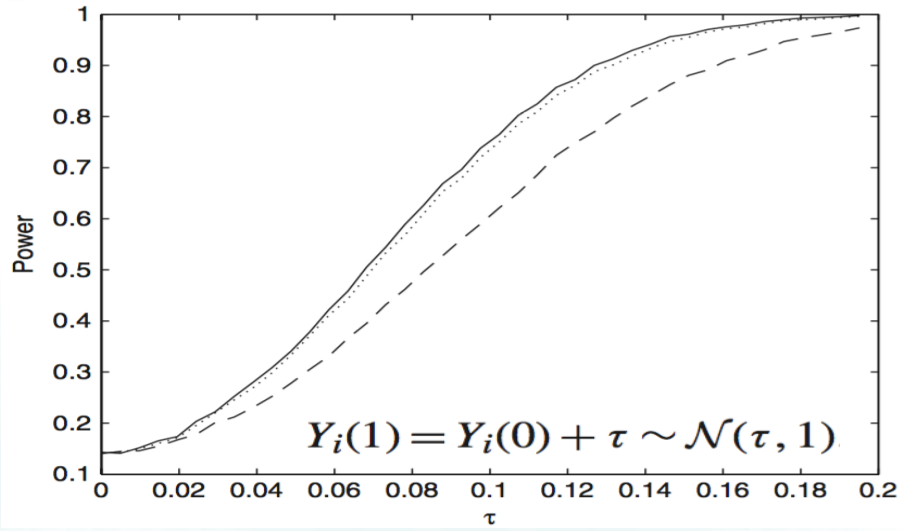
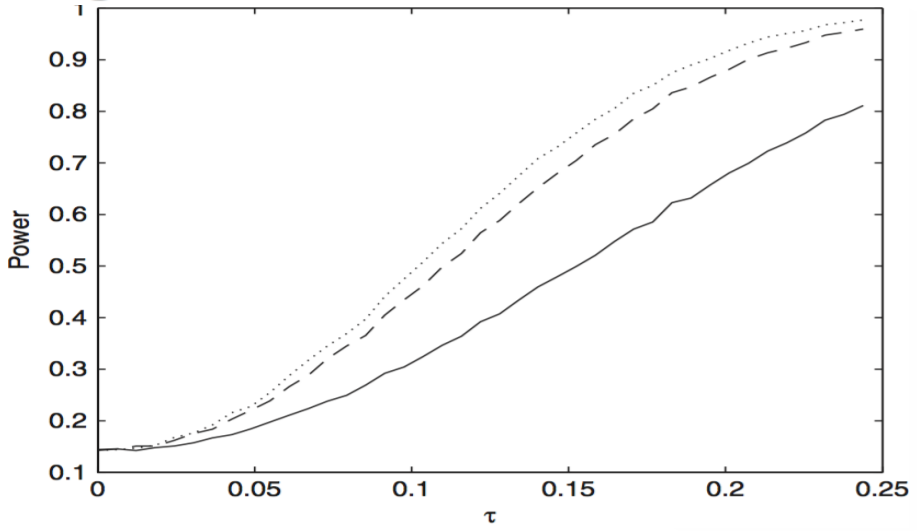


Figure 5.2. Additive model with outliers



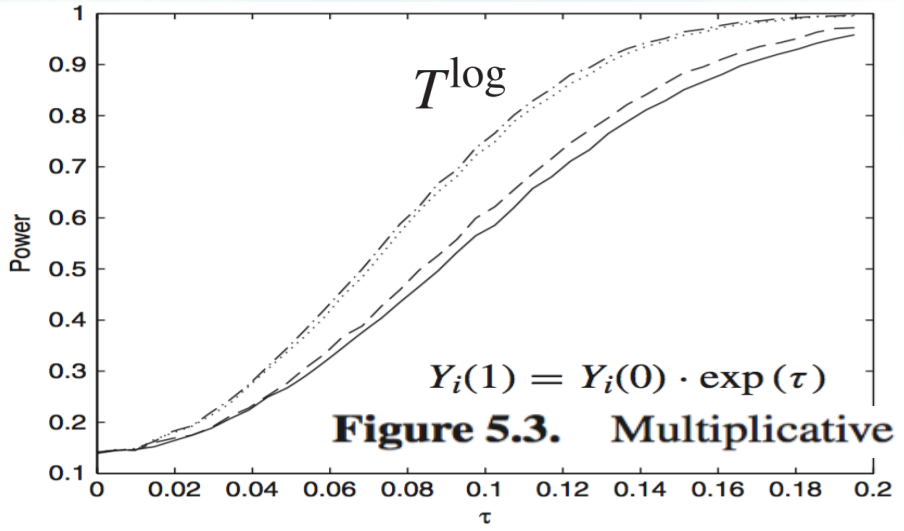


Figure 5.3. Multiplicative model

Conclusions

- rank-based statistic is an attractive choice: relatively good power & robust to different models
- the other choices for the test statistics performed well only in settings that play to their advantages, at the expense of relatively poor power in other settings.

Step 2. Choosing T (3): Statistics with Multiple Components

- Situations, e.g.
  - multiple alternative hypotheses of interest: e.g. different mean, different variance
  - two outcomes for each unit: e.g. Honey example
- Can we test each H_0 / use each T separately and conclude together based on multiple p-values?
- Possible solution:

$$T^{\text{comb}} = g(T^1, T^2),$$

$$T^{\text{Hotelling}} = \begin{pmatrix} \overline{Y}^{\text{obs}}_{\text{t},1} - \overline{Y}^{\text{obs}}_{\text{c},1} \\ \overline{Y}^{\text{obs}}_{\text{t},2} - \overline{Y}^{\text{obs}}_{\text{c},2} \end{pmatrix}' \left(\hat{V}_c/N_{\text{c}} + \hat{V}_t/N_{\text{t}} \right)^{-1} \begin{pmatrix} \overline{Y}^{\text{obs}}_{\text{t},1} - \overline{Y}^{\text{obs}}_{\text{c},1} \\ \overline{Y}^{\text{obs}}_{\text{t},2} - \overline{Y}^{\text{obs}}_{\text{c},2} \end{pmatrix}$$

Step 3. Computation of P-Values (1)

- Computing exact P-value is very expensive:

  Enumerate $\binom{N_c + N_t}{N_t}$ possible assignment vectors

- An precise approximation can be obtained by simulation
  - randomly choose a subset of K possible assignment vectors
  - For each assignment vector obtained, calculate its test statistics

$$T^{\text{dif},k} = \overline{Y}_{t,k} - \overline{Y}_{c,k}$$

  - approximate the p-value by

$$\hat{p} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{1}_{T^{\text{dif},k} \geq T^{\text{dif,obs}}}$$

CLT

large-sample standard error:

$$\sqrt{p^*(1-p^*)/K} \leq \frac{1}{2\sqrt{K}}$$

The true p-value

Step 3. Computation of P-Values (1)

Unit: $N = 72$

$$T^{\text{rank}} = |\bar{R}_t - \bar{R}_c| = \left| \frac{\sum_{i:W_i=1} R_i}{N_t} - \frac{\sum_{i:W_i=0} R_i}{N_c} \right|$$

Table 5.8. ***P-Values Estimated through Simulation for Honey Data from Table 5.1 for Null Hypothesis of Zero Effects***

Number of Simulations	P-Value	$\widehat{(\text{s.e.})}$
100	0.010	(0.010)
1,000	0.044	(0.006)
10,000	0.044	(0.002)
100,000	0.042	(0.001)
1,000,000	0.043	(0.000)



Step 3. Computation of P-Values (2)

- Under H_0, we have

$$\mathrm{E}\left(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}\right) = 0,$$

$$\mathrm{Var}\left(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}\right) = \frac{N}{N_t N_c} S^2,$$

where $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - \bar{Y}(1)\right)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0) - \bar{Y}(0)\right)^2$.

- According to the Finite Population CLT, under H_0, we have

$$\frac{\bar{Y}_t^{obs} - \bar{Y}_c^{obs}}{\sqrt{\frac{N}{N_t N_c} S^2}} \to^d N(0,1), \text{as } N \to +\infty$$

- Since S^2 is fixed under H_0, it is equivalent to use $\frac{\bar{Y}_t^{obs} - \bar{Y}_c^{obs}}{\sqrt{\frac{N}{N_t N_c} S^2}}$ and T^{dif} in FEP.

Reference for Finite Population CLT:

H´ ajek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematics Institute of the Hungarian Academy of Science*, 5:361–74.



Step 3. Computation of P-Values (2)

- According to the Finite Population CLT, under H_0, we have

$$\frac{\bar{Y}_t^{obs} - \bar{Y}_c^{obs}}{\sqrt{\frac{N}{N_t N_c} S^2}} \to^d N(0,1), \text{as } N \to +\infty$$

- Two-sample t-test assuming equal variance:

$$\frac{\bar{Y}_t^{obs} - \bar{Y}_c^{obs}}{\sqrt{\frac{N}{N_t N_c (N-2)} \left[\sum_{i:W_i=1} \left(Y_i(1) - \bar{Y}_t^{obs}\right)^2 + \sum_{i:W_i=0} \left(Y_i(0) - \bar{Y}_c^{obs}\right)^2\right]}} \sim t_{n-2}.$$

- For comparison, we have the expansion

$$(N-1)S^2 = \sum_{i:W_i=1} \left(Y_i(1) - \bar{Y}_t^{obs}\right)^2 + \sum_{i:W_i=0} \left(Y_i(0) - \bar{Y}_c^{obs}\right)^2 + \frac{N_t N_c}{N} \left(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}\right)^2.$$

- With a large N and under H_0, the approximate p-value from finite population CLT is close to the p-value from the classic two-sample t-test assuming equal variance, which can be calculated by t.test with var.equal = TRUE in R.

Step 3. Computation of P-Values (2)

- Similarly, according to the Finite Population CLT, under H_0, we have

$$\frac{\bar{Y}_t^{obs} - \bar{Y}_c^{obs}}{\sqrt{\frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}}} \to^d N(0,1), \text{as } N \to +\infty$$

- With a large N and under H_0, the approximate p-value from finite population CLT is close to the p-value from the classic two-sample t-test, which can be calculated by t.test with var.equal = FALSE in R.
- Similarly,
  - The function wilcox.test can compute both exact and asymptotic p-values based on the rank statistic;
  - The function ks.test can compute both exact and asymptotic p-values based on the Kolmogorov-Smirnov Statistic.

Thinking: A Simple Example with 6 Units

Table 5.3. ***Cough Frequency for the First Six Units from the Honey Study***

Unit	Potential Outcomes				
	Cough Frequency (cfa)		Observed Variables		
	$Y_i(0)$	$Y_i(1)$	W_i	X_i (cfp)	Y_i^{obs} (cfa)
1	?	3	1	4	3
2	?	5	1	6	5
3	?	0	1	4	0
4	4	?	0	4	4
5	0	?	0	1	0
6	1	?	0	5	1

Results are consistent with intuition?

Discussion 1: FEP with Covariates

- Gain score:

$$Y_i'(w) = Y_i(w) - X_i,$$

$$Y_i'^{,\text{obs}} = Y_i^{\text{obs}} - X_i = \begin{cases} Y_i'(0) & \text{if } W_i = 0, \\ Y_i'(1) & \text{if } W_i = 1. \end{cases}$$

$$\begin{aligned} T^{\text{gain}} &= \frac{\sum_{i:W_i=1} Y_i'^{,\text{obs}}}{N_{\text{t}}} - \frac{\sum_{i:W_i=0} Y_i'^{,\text{obs}}}{N_{\text{c}}} \\ &= \frac{\sum_{i:W_i=1} (Y_i^{\text{obs}} - X_i)}{N_{\text{t}}} - \frac{\sum_{i:W_i=0} (Y_i^{\text{obs}} - X_i)}{N_{\text{c}}} \\ &= \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - (\overline{X}_t - \overline{X}_c), \end{aligned}$$

VS

$$T^{\text{dif}} = \left| \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} \right|$$

0 on average (average in what sense?), but not 0 for each time

Discussion 1: FEP with Covariates

- Proportional change:

$$Y_i''(w) = \frac{Y_i(w) - X_i}{X_i}, \quad \text{for } w = 0, 1, \longrightarrow \frac{Y_i(1) - X_i}{X_i} - \frac{Y_i(0) - X_i}{X_i} = \frac{Y_i(1) - Y_i(0)}{X_i}$$

$$Y_i''^{,\text{obs}} = \frac{Y_i^{\text{obs}} - X_i}{X_i}.$$

$$T^{\text{prop-change}} = \overline{Y''}_t - \overline{Y''}_c = \frac{1}{N_t} \sum_{i:W_i=1} \frac{Y_i^{\text{obs}} - X_i}{X_i} - \frac{1}{N_c} \sum_{i:W_i=0} \frac{Y_i^{\text{obs}} - X_i}{X_i}.$$

- Regression coefficient:

$$\left(\hat{\beta}_0, \hat{\beta}_X, \hat{\beta}_W\right) = \arg \min_{\beta_0, \beta_X, \beta_W} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \beta_0 - \beta_X \cdot X_i - \beta_W \cdot W_i\right)^2$$

$$T^{\text{reg-coef}} = \hat{\beta}_W.$$

- What if the model is wrong?
- Is it still valid?
- Is it more effective?
- What is the causal meaning for the coefficient?

Discussion 1: The Honey Experiment Data

Unit: $N = 72$
Calculation of p-values: 1,000,000 draws

$p = P(T \geq T^{obs})$

This reflects the strong correlation between the prior cough frequency and *ex post* cough frequency

Table 5.9. ***P-Values for Honey Data from Table 5.1, for Null Hypothesis of Zero Effects Using Various Statistics***

Test Statistic	Statistic	P-Value
T^{dif}	−0.697	0.067
T^{quant} $(\delta = 0.25)$	−1.000	0.440
T^{quant} $(\delta = 0.50)$	−1.000	0.637
T^{quant} $(\delta = 0.75)$	−1.000	0.576
$T^{\text{t-stat}}$	−1.869	0.065
T^{rank}	−9.785	0.043
T^{ks}	0.304	0.021
$T^{\text{F-stat}}$	3.499	0.182
T^{gain}	−0.967	0.006
$T^{\text{reg-coef}}$	−0.911	0.008

Discussion 2: Fisherian / Fiducial Interval

Table 5.7. ***P-Values for Tests of Constant Treatment Effects (Full Honey Data Set from Table 5.1, with Cough Frequency as Outcome)***

Hypothesized Treatment Effect	P-Value (level) T^{dif}	P-Value (rank) T^{rank}
−3.00	0.000	0.000
−2.75	0.000	0.000
−2.50	0.000	0.000
−2.25	0.000	0.000
−2.00	0.001	0.000
−1.75	0.006	0.078
−1.50	0.037	0.078
−1.44	0.050	0.078
−1.25	0.146	0.078
−1.00	0.459	0.628
−0.75	0.897	0.428
−0.50	0.604	0.428
−0.25	0.237	0.429
0.00	0.067	0.043
0.06	0.050	0.043
0.25	0.014	0.001
0.50	0.003	0.000
0.75	0.000	0.001
1.00	0.000	0.000

$$H_0 : Y_i(1) = Y_i(0) + C, \text{ for } i = 1, \ldots, N.$$

$$p = P(T \geq T^{obs})$$

95% “Fisher” interval for the treatment effect

Neyman's approach

HOW TO ESTIMATE THE AVERAGE CAUSAL EFFECT?

FINITE SAMPLE



Fisher's Approach vs Neyman's Approach

- Fill in missing potential outcomes based on

$$H_0 : Y_i(0) = Y_i(1), \quad \text{for } i = 1, \ldots, N.$$

- Limitations:
  - The sharp null hypothesis is somewhat restrictive
  - Rejecting the null hypothesis itself cannot provide us an estimator of the unknown (average) treatment effect
- Neyman's repeated sampling approach
  - Focus on average causal effect instead of unit-level effect
  - Provide unbiased point estimator & interval estimator
  - Randomness: considering only finite population first
    - randomized assignment of treatments to the units in the sample



Estimation of Average Treatment Effect

- Causal **estimand**:

$$\tau_{\mathrm{fs}} = \frac{1}{N}\sum_{i=1}^{N}\big(Y_i(1) - Y_i(0)\big) = \overline{Y}(1) - \overline{Y}(0),$$

$$\overline{Y}(0) = \frac{1}{N}\sum_{i=1}^{N} Y_i(0), \quad \text{and} \quad \overline{Y}(1) = \frac{1}{N}\sum_{i=1}^{N} Y_i(1).$$

- **Estimator**:

$$\hat{\tau}^{\mathrm{dif}} = \overline{Y}_{\mathrm{t}}^{\mathrm{obs}} - \overline{Y}_{\mathrm{c}}^{\mathrm{obs}},$$

How to express it in terms of $Y_i(1)$, $Y_i(0)$, W

$$\overline{Y}_{\mathrm{c}}^{\mathrm{obs}} = \frac{1}{N_{\mathrm{c}}}\sum_{i:W_i=0} Y_i^{\mathrm{obs}} \quad \text{and} \quad \overline{Y}_{\mathrm{t}}^{\mathrm{obs}} = \frac{1}{N_{\mathrm{t}}}\sum_{i:W_i=1} Y_i^{\mathrm{obs}}.$$

Unbiasness of the Estimator

$$\hat{\tau}^{\text{dif}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{W_i \cdot Y_i(1)}{N_t/N} - \frac{(1-W_i)\cdot Y_i(0)}{N_c/N}\right)$$

$$\mathbb{E}_W[W_i|\mathbf{Y}(0),\mathbf{Y}(1)] = \Pr_W(W_i = 1|\mathbf{Y}(0),\mathbf{Y}(1)) = N_t/N$$

$$\mathbb{E}_W\left[\hat{\tau}^{\text{dif}}\middle|\mathbf{Y}(0),\mathbf{Y}(1)\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbb{E}_W[W_i]\cdot Y_i(1)}{N_t/N} - \frac{\mathbb{E}_W[1-W_i])\cdot Y_i(0)}{N_c/N}\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\big(Y_i(1) - Y_i(0)\big) = \tau_{\text{fs}}.$$

Example: The Honey Experiment Data

Unit: $N = 72$ children
Treatment: a single dose of buckwheat honey
Control: no active treatment
Two outcomes:
cough frequency afterwards (cfa) & cough severity afterwards (csa)
Two covariates:
cough frequency prior (cfp) & cough severity prior (csp)
Completely randomized assignment:
$N_t = 35$ & $N_c = 37$

$$\hat{\tau}^{\text{dif}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} = -0.697$$



Steps for Building Confidence Interval

- 3 steps:
  - derive the sampling variance of the estimator for the average treatment effect
  - develop estimators for this sampling variance
  - create a large-sample confidence interval for the average treatment effect



The Sampling Variance of the Neyman Estimator with N Units (I)

$$\mathbb{E}_W\left[W_i^2\right] = \mathbb{E}_W\left[W_i\right] = \frac{N_{\mathrm{t}}}{N}, \quad \text{and} \quad \mathbb{V}_W(W_i) = \frac{N_{\mathrm{t}}}{N}\cdot\left(1-\frac{N_{\mathrm{t}}}{N}\right).$$

$$\mathbb{E}_W[W_i\cdot W_{i'}] = \Pr_W(W_i=1)\cdot\Pr_W(W_{i'}=1|W_i=1) = \frac{N_{\mathrm{t}}}{N}\cdot\frac{N_{\mathrm{t}}-1}{N-1}, \quad \text{for } i\neq j,$$

Theorem 6.2 *The sampling variance of* $\hat{\tau}^{\mathrm{dif}} = \overline{Y}_{\mathrm{t}}^{\mathrm{obs}} - \overline{Y}_{\mathrm{c}}^{\mathrm{obs}}$ *is*

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}} - \overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right) = \frac{S_c^2}{N_{\mathrm{c}}} + \frac{S_t^2}{N_{\mathrm{t}}} - \frac{S_{tc}^2}{N},$$

variance of $Y_i(0)$ in the finite population

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0)-\overline{Y}(0)\right)^2,$$

$$S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)^2,$$

variance of $Y_i(1)$ in the finite population

variance of the unit-level treatment effects in the finite population

$$S_{tc}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-Y_i(0)-(\overline{Y}(1)-\overline{Y}(0))\right)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-Y_i(0)-\tau_{\mathrm{fs}}\right)^2.$$

Key Steps of the Proof

$$\hat{\tau}^{\text{dif}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} = \frac{1}{N_{\text{t}}} \sum_{i=1}^{N} W_i \cdot Y_i^{\text{obs}} - \frac{1}{N_{\text{c}}} \sum_{i=1}^{N} (1 - W_i) \cdot Y_i^{\text{obs}}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{N}{N_{\text{t}}} \cdot W_i \cdot Y_i(1) - \frac{N}{N_{\text{c}}} \cdot (1 - W_i) \cdot Y_i(0) \right)$$

$$D_i = W_i - \frac{N_{\text{t}}}{N} = \begin{cases} \frac{N_{\text{c}}}{N} & \text{if } W_i = 1 \\ -\frac{N_{\text{t}}}{N} & \text{if } W_i = 0. \end{cases}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{N}{N_{\text{t}}} \cdot \left(D_i + \frac{N_{\text{t}}}{N} \right) \cdot Y_i(1) - \frac{N}{N_{\text{c}}} \cdot \left(\frac{N_{\text{c}}}{N} - D_i \right) \cdot Y_i(0) \right)$$

$$= \frac{1}{N} \sum_{i=1}^{N} (Y_i(1) - Y_i(0)) + \frac{1}{N} \sum_{i=1}^{N} D_i \cdot \left(\frac{N}{N_{\text{t}}} \cdot Y_i(1) + \frac{N}{N_{\text{c}}} \cdot Y_i(0) \right)$$

$$= \tau_{fs} + \frac{1}{N} \sum_{i=1}^{N} D_i \cdot \left(\frac{N}{N_{\text{t}}} \cdot Y_i(1) + \frac{N}{N_{\text{c}}} \cdot Y_i(0) \right).$$

$$Y_i^+$$

$$\mathbb{V}_W \left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} \right) = \mathbb{V}_W \left(\frac{1}{N} \sum_{i=1}^{N} D_i \cdot Y_i^+ \right) = \frac{1}{N^2} \cdot \mathbb{E}_W \left[\left(\sum_{i=1}^{N} D_i \cdot Y_i^+ \right)^2 \right]$$



Key Steps of the Proof

$$D_i = W_i - \frac{N_t}{N} = \begin{cases} \frac{N_c}{N} & \text{if } W_i = 1 \\ -\frac{N_t}{N} & \text{if } W_i = 0. \end{cases}$$

$$\Pr_W\left(D_i \cdot D_j = d\right) = \begin{cases} \frac{N_t \cdot (N_t - 1)}{N \cdot (N-1)} & \text{if } d = N_c^2/N^2 \\ 2 \cdot \frac{N_t \cdot N_c}{N \cdot (N-1)} & \text{if } d = -N_t N_c/N^2 \\ \frac{N_c \cdot (N_c - 1)}{N \cdot (N-1)} & \text{if } d = N_t^2/N^2 \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbb{E}_W[D] = 0,$$

$$\mathbb{V}(D_i) = \mathbb{E}[D_i^2] = N_c N_t / N^2$$

$$\mathbb{E}_W\left[D_i \cdot D_j\right] = \begin{cases} \frac{N_c \cdot N_t}{N^2} & \text{if } i = j \\ -\frac{N_t \cdot N_c}{N^2 \cdot (N-1)} & \text{if } i \neq j \end{cases}.$$

The Sampling Variance (I): A Special Case

Theorem 6.2 *The sampling variance of* $\hat{\tau}^{\text{dif}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}$ *is*

$$\mathbb{V}_W\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right) = \frac{S_c^2}{N_{\text{c}}} + \frac{S_t^2}{N_{\text{t}}} - \frac{S_{tc}^2}{N},$$

variance of $Y_i(0)$ *in the finite population*

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0) - \overline{Y}(0)\right)^2,$$

$$S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - \overline{Y}(1)\right)^2,$$

variance of $Y_i(1)$ *in the finite population*

variance of the unit-level treatment effects in the finite population

$$S_{tc}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0) - (\overline{Y}(1) - \overline{Y}(0))\right)^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0) - \tau_{\text{fs}}\right)^2.$$

If $Y_i(1) - Y_i(0) = \tau_{fs}, \forall i,$ then we have $S^2{}_{tc} = 0$, and thus

$$\mathbb{V}_W\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right) = \mathbb{V}^{\text{const}} = \frac{S_c^2}{N_{\text{c}}} + \frac{S_t^2}{N_{\text{t}}}$$



The Sampling Variance (II): Alternative Representation

$$S_{tc}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0) - (\overline{Y}(1) - \overline{Y}(0))\right)^2$$

$$S_{tc}^2 = S_c^2 + S_t^2 - 2 \cdot \rho_{tc} \cdot S_c \cdot S_t$$

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0) - \overline{Y}(0)\right)^2, \quad S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - \overline{Y}(1)\right)^2,$$

$$\rho_{tc} = \frac{1}{(N-1)\cdot S_c \cdot S_t}\sum_{i=1}^{N}\left(Y_i(1) - \overline{Y}(1)\right)\cdot\left(Y_i(0)\right) - \overline{Y}(0))$$

the correlation coefficient

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}} - \overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right) = \frac{S_c^2}{N_{\mathrm{c}}} + \frac{S_t^2}{N_{\mathrm{t}}} - \frac{S_{tc}^2}{N}$$

$$= \frac{N_{\mathrm{t}}}{N \cdot N_{\mathrm{c}}} \cdot S_c^2 + \frac{N_{\mathrm{c}}}{N \cdot N_{\mathrm{t}}} \cdot S_t^2 + \frac{2}{N} \cdot \rho_{tc} \cdot S_c \cdot S_t$$

The Sampling Variance (II): Alternative Representation

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}}-\overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right)=\frac{N_{\mathrm{t}}}{N\cdot N_{\mathrm{c}}}\cdot S_c^2+\frac{N_{\mathrm{c}}}{N\cdot N_{\mathrm{t}}}\cdot S_t^2+\frac{2}{N}\cdot\rho_{tc}\cdot S_c\cdot S_t$$

$$S_c^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0)-\overline{Y}(0)\right)^2,\qquad S_t^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)^2,$$

correlation coefficient $\rho_{tc}=\dfrac{1}{(N-1)\cdot S_c\cdot S_t}\displaystyle\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)\cdot\left(Y_i(0))-\overline{Y}(0)\right)$

- Reach minimum when $\rho_{tc}=-1$

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}}-\overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\middle|\,\rho_{tc}=-1\right)=\frac{N_{\mathrm{t}}}{N\cdot N_{\mathrm{c}}}\cdot S_c^2+\frac{N_{\mathrm{c}}}{N\cdot N_{\mathrm{t}}}\cdot S_t^2-\frac{2}{N}\cdot S_c\cdot S_t$$

- Reach maximum when $\rho_{tc}=1$

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}}-\overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\middle|\,\rho_{tc}=1\right)=\frac{N_{\mathrm{t}}}{N\cdot N_{\mathrm{c}}}\cdot S_c^2+\frac{N_{\mathrm{c}}}{N\cdot N_{\mathrm{t}}}\cdot S_t^2+\frac{2}{N}\cdot S_c\cdot S_t=\frac{S_c^2}{N_{\mathrm{c}}}+\frac{S_t^2}{N_{\mathrm{t}}}-\frac{(S_c-S_t)^2}{N}.$$

Summary for $\mathrm{Var}(\hat{\tau}^{dif})$

- Without additional assumptions, we have

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}} - \overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right) = \frac{S_c^2}{N_{\mathrm{c}}} + \frac{S_t^2}{N_{\mathrm{t}}} - \frac{S_{tc}^2}{N}$$

表达方式一

$$= \frac{N_{\mathrm{t}}}{N \cdot N_{\mathrm{c}}} \cdot S_c^2 + \frac{N_{\mathrm{c}}}{N \cdot N_{\mathrm{t}}} \cdot S_t^2 + \frac{2}{N} \cdot \rho_{tc} \cdot S_c \cdot S_t$$

表达方式二

- With an assumption: $Y_i(1) - Y_i(0) = \tau_{fs}, \forall i,$ we have

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}} - \overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right) = \mathbb{V}^{\mathrm{const}} = \frac{S_c^2}{N_{\mathrm{c}}} + \frac{S_t^2}{N_{\mathrm{t}}}$$

Estimating the Sample Variance (I)

Theorem 6.3 *If the treatment effect $Y_i(1) - Y_i(0)$ is constant, then an unbiased estimator for the sampling variance is*

$$\hat{\mathbb{V}}^{\text{neyman}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}.$$

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i(0) - \overline{Y}_c^{\text{obs}}\right)^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right)^2$$

$$s_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i(1) - \overline{Y}_t^{\text{obs}}\right)^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}}\right)^2$$

If $S_t^2 = S_c^2$

$$\hat{\mathbb{V}}^{\text{const}} = s^2 \cdot \left(\frac{1}{N_c} + \frac{1}{N_t}\right)$$

$$s^2 = \frac{1}{N - 2} \cdot \left(s_c^2 \cdot (N_c - 1) + s_t^2 \cdot (N_t - 1)\right)$$

$$= \frac{1}{N - 2} \cdot \left(\sum_{i:W_i=0} \left(Y_i^{\text{obs}} - \overline{Y}_c^{\text{obs}}\right)^2 + \sum_{i:W_i=1} \left(Y_i^{\text{obs}} - \overline{Y}_t^{\text{obs}}\right)^2 \right)$$

Estimating the Sample Variance (II)

A pre-specified ρ_{tc} leads to the estimate below:

$$\hat{\mathbb{V}}^{\rho_{tc}} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_t^2 \cdot \frac{N_c}{N \cdot N_t} + \rho_{tc} \cdot s_c \cdot s_t \cdot \frac{2}{N}$$

$$s_c^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i(0) - \overline{Y}_c^{obs}\right)^2 = \frac{1}{N_c - 1} \sum_{i:W_i=0} \left(Y_i^{obs} - \overline{Y}_c^{obs}\right)^2$$

$$s_t^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i(1) - \overline{Y}_t^{obs}\right)^2 = \frac{1}{N_t - 1} \sum_{i:W_i=1} \left(Y_i^{obs} - \overline{Y}_t^{obs}\right)^2$$

Provide a proper **up-bound**

Specify $\rho_{tc} = 1$

$$\hat{\mathbb{V}}^{\rho_{tc}=1} = s_c^2 \cdot \frac{N_t}{N \cdot N_c} + s_1^2 \cdot \frac{N_c}{N \cdot N_t} + s_c \cdot s_t \cdot \frac{2}{N}$$

$$= \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t} - \frac{(s_t - s_c)^2}{N}.$$

Summary for $\widehat{\mathrm{Var}}(\hat{\tau}^{dif})$

- Without additional assumptions

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}}-\overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right)=\frac{S_c^2}{N_{\mathrm{c}}}+\frac{S_t^2}{N_{\mathrm{t}}}-\frac{S_{tc}^2}{N} \xrightarrow{\text{conservative}} \hat{\mathbb{V}}^{\mathrm{neyman}}=\frac{s_c^2}{N_{\mathrm{c}}}+\frac{s_t^2}{N_{\mathrm{t}}}.$$ 表达方式一

$$=\frac{N_{\mathrm{t}}}{N\cdot N_{\mathrm{c}}}\cdot S_c^2+\frac{N_{\mathrm{c}}}{N\cdot N_{\mathrm{t}}}\cdot S_t^2+\frac{2}{N}\cdot\rho_{tc}\cdot S_c\cdot S_t$$ 表达方式二

conservative →

$$\hat{\mathbb{V}}^{\rho_{tc}=1}=s_c^2\cdot\frac{N_{\mathrm{t}}}{N\cdot N_{\mathrm{c}}}+s_1^2\cdot\frac{N_{\mathrm{c}}}{N\cdot N_{\mathrm{t}}}+s_c\cdot s_t\cdot\frac{2}{N}$$
$$=\frac{s_c^2}{N_{\mathrm{c}}}+\frac{s_t^2}{N_{\mathrm{t}}}-\frac{(s_t-s_c)^2}{N}.$$

- With an assumption $Y_i(1)-Y_i(0)=\tau_{fs},\forall i,$

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}}-\overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right)=\mathbb{V}^{\mathrm{const}}=\frac{S_c^2}{N_{\mathrm{c}}}+\frac{S_t^2}{N_{\mathrm{t}}} \xrightarrow{\text{unbiased}} \hat{\mathbb{V}}^{\mathrm{const}}=s^2\cdot\left(\frac{1}{N_{\mathrm{c}}}+\frac{1}{N_{\mathrm{t}}}\right)$$

Confidence Intervals

Based on a normal approximation to the randomization distribution of $\hat{\tau}^{dif}$

$$\mathrm{CI}^{1-\alpha}(\tau_{\mathrm{fs}}) = \left(\hat{\tau}^{\mathrm{dif}} + z_{\alpha/2} \cdot \sqrt{\hat{\mathbb{V}}}, \hat{\tau}^{\mathrm{dif}} + z_{1-\alpha/2} \cdot \sqrt{\hat{\mathbb{V}}}\right)$$

an estimate of the sampling variance of τ^{dif} , e.g.,

$$\hat{\mathbb{V}}^{\mathrm{neyman}} = \frac{s_c^2}{N_{\mathrm{c}}} + \frac{s_t^2}{N_{\mathrm{t}}} \qquad \hat{\mathbb{V}}^{\mathrm{const}} = s^2 \cdot \left(\frac{1}{N_{\mathrm{c}}} + \frac{1}{N_{\mathrm{t}}}\right)$$

$$\hat{\mathbb{V}}^{\rho_{tc}=1} = s_c^2 \cdot \frac{N_{\mathrm{t}}}{N \cdot N_{\mathrm{c}}} + s_t^2 \cdot \frac{N_{\mathrm{c}}}{N \cdot N_{\mathrm{t}}} + s_c \cdot s_t \cdot \frac{2}{N}$$

- E.g., confidence interval when $\alpha = 0.10$

$$\mathrm{CI}^{0.90}(\tau_{\mathrm{fs}}) = \left(\hat{\tau}^{\mathrm{dif}} - 1.645 \cdot \sqrt{\hat{\mathbb{V}}}, \hat{\tau}^{\mathrm{dif}} + 1.645 \cdot \sqrt{\hat{\mathbb{V}}}\right)$$

Testing

Neyman approach

$$H_0^{\text{neyman}}: \quad \frac{1}{N}\sum_{i=1}^{N}(Y_i(1)-Y_i(0))=0, \text{ and}$$

$$H_a^{\text{neyman}}: \quad \frac{1}{N}\sum_{i=1}^{N}(Y_i(1)-Y_i(0))\neq 0.$$

T-statistics: $$t=\frac{\overline{Y}_t^{\text{obs}}-\overline{Y}_c^{\text{obs}}}{\sqrt{\hat{\mathbb{V}}^{\text{neyman}}}}$$

Fisher approach (Exact P-Value)

$$H_0^{\text{fisher}}: \quad Y_i(1)-Y_i(0)=0 \text{ for all } i=1,\ldots,N,$$

$$H_a^{\text{fisher}}: \quad Y_i(1)-Y_i(0)\neq 0 \text{ for some } i=1,\ldots,N.$$

Test statistics: $$T^{\text{dif}}=\left|\overline{Y}_{\text{t}}^{\text{obs}}-\overline{Y}_{\text{c}}^{\text{obs}}\right|$$



Example: The Honey Experiment Data

Unit: $N = 72$ children
Treatment: a single dose of buckwheat honey
Control: no active treatment
Two outcomes:
cough frequency afterwards (cfa)
Completely randomized assignment:
$N_t = 35$ & $N_c = 37$

Approach	*P*-value
Neyman's	0.062
Fisher's	0.067

$$\hat{\tau}^{\text{dif}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} = -0.697$$

$$\hat{\mathbb{V}}^{\text{neyman}} = \frac{s_c^2}{N_{\text{c}}} + \frac{s_t^2}{N_{\text{t}}} \approx 0.139 \qquad \hat{\mathbb{V}}^{\rho_{tc}=1} = \frac{s_c^2}{N_{\text{c}}} + \frac{s_t^2}{N_{\text{t}}} - \frac{(s_t - s_c)^2}{N} \approx 0.139$$

$$\hat{\mathbb{V}}^{\text{const}} = s^2 \cdot \left(\frac{1}{N_{\text{c}}} + \frac{1}{N_{\text{t}}}\right) \approx 0.139$$

$$\text{CI}^{0.90}(\tau_{\text{fs}}) = \left(\hat{\tau}^{\text{dif}} - 1.645 \cdot \sqrt{\hat{\mathbb{V}}}, \hat{\tau}^{\text{dif}} + 1.645 \cdot \sqrt{\hat{\mathbb{V}}}\right) \approx [-1.31, -0.08]$$

Neyman's approach

HOW TO ESTIMATE THE AVERAGE CAUSAL EFFECT?

SUPER POPULATION



Fisher's Approach vs Neyman's Approach

- Fill in missing potential outcomes based on

$$H_0 : Y_i(0) = Y_i(1), \quad \text{for } i = 1, \ldots, N.$$

- Limitations:
  - The sharp null hypothesis is somewhat restrictive
  - Rejecting the null hypothesis itself cannot provide us an estimator of the unknown (average) treatment effect
- Neyman's repeated sampling approach
  - Focus on average causal effect instead of unit-level effect
  - Provide unbiased point estimator & interval estimator
  - Consider two layers of randomness:
    - repeated sampling from the population
    - randomized assignment of treatments to the units in the sample



Super Population Average Treatment Effects

$$\tau_{\text{fs}} = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) \quad \dashrightarrow \quad \tau_{\text{sp}} = \mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0)\right]$$

$$\mathbb{E}_{\text{sp}}\left[\tau_{\text{fs}}\right] = \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0)\right] = \tau_{\text{sp}}$$

$$\sigma_c^2 = \mathbb{V}_{\text{sp}}(Y_i(0)) = \mathbb{E}_{\text{sp}}\left[(Y_i(0) - \mathbb{E}_{\text{sp}}[Y_i(0))^2\right]$$

$$\sigma_t^2 = \mathbb{V}_{\text{sp}}(Y_i(1)) = \mathbb{E}_{\text{sp}}\left[(Y_i(1) - \mathbb{E}_{\text{sp}}[Y_i(1))^2\right]$$

$$\sigma_{tc}^2 = \mathbb{V}_{\text{sp}}(Y_i(1) - Y_i(0)) = \mathbb{E}_{\text{sp}}[(Y_i(1) - Y_i(0) - \tau_{\text{sp}})^2]$$

$$\mathbb{V}_{\text{sp}}(\tau_{\text{fs}}) = \mathbb{V}_{\text{sp}}\left(\overline{Y}(1) - \overline{Y}(0)\right) = \sigma_{tc}^2/N$$



Inference for Population Average Treatment Effects

- Is the estimator $\hat{\tau}^{dif}$ unbiased?
- To build the corresponding confidence interval, what should we get?

$$\mathbb{V}_{\text{sp}} = \mathbb{V}_{\text{sp}}\left(\hat{\tau}^{\text{dif}}\right) = \quad ?$$



Inference for Population ATE

$$\mathbb{V}\left(\hat{\tau}^{\text{dif}}\right) = \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \mathbb{E}\left[\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right) + \left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right)^2\right] \dashrightarrow \frac{\sigma_c^2}{N_{\text{c}}} + \frac{\sigma_t^2}{N_{\text{t}}} - \frac{\sigma_{tc}^2}{N}$$

$$+ \mathbb{E}_{\text{sp}}\left[\left(\left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right] \dashrightarrow \mathbb{V}_{\text{sp}}(\tau_{\text{fs}}) = \mathbb{V}_{\text{sp}}\left(\overline{Y}(1) - \overline{Y}(0)\right) = \sigma_{tc}^2/N$$

$$+ 2 \cdot \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right) \cdot \left(\left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)\right]$$

$$\mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \overline{Y}(1) - \overline{Y}(0)\right)^2\right]$$

$$= \mathbb{E}_{\text{sp}}\left[\mathbb{E}_W\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \overline{Y}(1) - \overline{Y}(0)\right)^2 \middle| \mathbf{Y}(0), \mathbf{Y}(1)\right]\right]$$

Based on Theorem 6.2

$$= \mathbb{E}_{\text{sp}}\left[\frac{S_c^2}{N_{\text{c}}} + \frac{S_t^2}{N_{\text{t}}} - \frac{S_{tc}^2}{N}\right] = \frac{\sigma_c^2}{N_{\text{c}}} + \frac{\sigma_t^2}{N_{\text{t}}} - \frac{\sigma_{tc}^2}{N}.$$

Inference for Population ATE

$$\mathbb{V}\left(\hat{\tau}^{\text{dif}}\right) = \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \mathbb{E}\left[\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right) + \left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right]$$

$$= \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right)^2\right] \dashrightarrow \frac{\sigma_c^2}{N_{\text{c}}} + \frac{\sigma_t^2}{N_{\text{t}}} - \frac{\sigma_{tc}^2}{N}$$

$$+ \mathbb{E}_{\text{sp}}\left[\left(\left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)^2\right] \dashrightarrow \frac{1}{N}\mathbb{E}_{\text{sp}}[(Y_i(1) - Y_i(0) - \tau_{\text{sp}})^2] = \frac{1}{N}\sigma_{tc}^2$$

$$+ 2 \cdot \mathbb{E}\left[\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right) \cdot \left(\left(\overline{Y}(1) - \overline{Y}(0)\right) - \mathbb{E}_{\text{sp}}\left[\overline{Y}(1) - \overline{Y}(0)\right]\right)\right] \dashrightarrow 0$$

$$\mathbb{V}_{\text{sp}} = \mathbb{V}_{\text{sp}}\left(\hat{\tau}^{\text{dif}}\right) = \frac{\sigma_c^2}{N_{\text{c}}} + \frac{\sigma_t^2}{N_{\text{t}}} \Rightarrow \hat{\mathbb{V}}^{\text{sp}} = \frac{s_c^2}{N_{\text{c}}} + \frac{s_t^2}{N_{\text{t}}}$$

Summary



Summary

- FEP and Neyman's approach have different inference goals.
- FEP
  - Hypothesis testing only, no average causal effect estimation
  - Exact null distribution: approximation by simulation
  - Statistical power is the standard for choosing statistics
- Neyman's Approach
  - Finite sample
    - Unbiased estimator
    - Confidence interval: valid, effective
  - Super population: two layers of randomness

Limitations: how to incorporate covariates?

Reference

- Reference: [T1] Chapter 5, 6

Appendix: Proof of the sampling variance $\mathrm{Var}(\hat{\tau}^{dif})$



The Sampling Variance (I)

Theorem 6.2 *The sampling variance of* $\hat{\tau}^{\text{dif}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}$ *is*

$$\mathbb{V}_W\left(\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right) = \frac{S_c^2}{N_{\text{c}}} + \frac{S_t^2}{N_{\text{t}}} - \frac{S_{tc}^2}{N},$$

variance of $Y_i(0)$ *in the finite population*

variance of the unit-level treatment effects in the finite population

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0) - \overline{Y}(0)\right)^2, \qquad S_{tc}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0) - (\overline{Y}(1) - \overline{Y}(0))\right)^2$$

$$S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - \overline{Y}(1)\right)^2, \qquad = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0) - \tau_{\text{fs}}\right)^2.$$

variance of $Y_i(1)$ *in the finite population*

Proof for (I)

谢谢刘开星同学的分享!

$$V_w(\bar{Y}_t^{obs} - \bar{Y}_c^{obs}) = V_w\left(\frac{1}{N}\sum_i D_i Y_i^+\right) = \frac{1}{N^2} E_w[(\sum_i D_i Y_i^+)^2]$$

$$= \frac{1}{N^2}\left[\sum_i Y_i^{+2} E_w(D_i^2) + \sum_i \sum_{j\neq i} Y_i^+ Y_j^+ E_w(D_i D_j)\right]$$

$$= \frac{1}{N^2}\left[\frac{N_c N_t}{N^2}\sum_i \left(\frac{N}{N_t}Y_i(1) + \frac{N}{N_c}Y_i(0)\right)^2 - \frac{N_c N_t}{N^2(N-1)}\sum_i \sum_{j\neq i}\left(\frac{N}{N_t}Y_i(1) + \frac{N}{N_c}Y_i(0)\right)\left(\frac{N}{N_t}Y_j(1) + \frac{N}{N_c}Y_j(0)\right)\right]$$

$$= \frac{1}{N^2}\left[N_c N_t\left(\frac{\sum_i Y_i(1)^2}{N_t^2} + \frac{2\sum_i Y_i(1)Y_i(0)}{N_t N_c} + \frac{\sum_i Y_i(0)^2}{N_c^2}\right) - \frac{N_c N_t}{N-1}\sum_i \left(\frac{Y_i(1)}{N_t} + \frac{Y_i(0)}{N_c}\right)\left(\frac{N\bar{Y}(1) - Y_i(1)}{N_t} + \frac{N\bar{Y}(0) - Y_i(0)}{N_c}\right)\right]$$

$$= \frac{1}{N^2}\left[\left(\frac{N_c}{N_t}\sum_i Y_i^2(1) + 2\sum_i Y_i(1)Y_i(0) + \frac{N_t}{N_c}\sum_i Y_i(0)^2\right) - \frac{N_c N_t}{N-1}\sum_i \left(\frac{N\bar{Y}(1)Y_i(1) - Y_i(1)^2}{N_t^2} + \frac{N\bar{Y}(0)Y_i(1) - Y_i(0)Y_i(1)}{N_t N_c} + \frac{N\bar{Y}(1)Y_i(0) - Y_i(0)Y_i(1)}{N_t N_c} + \frac{N\bar{Y}(0)Y_i(0) - Y_i(0)^2}{N_c^2}\right)\right]$$

$$= \frac{1}{N^2}\left[\left(\frac{N_c}{N_t}\sum_i Y_i^2(1) + 2\sum_i Y_i(1)Y_i(0) + \frac{N_t}{N_c}\sum_i Y_i(0)^2\right) - \frac{N_c N_t}{N-1}\left(\frac{N^2\bar{Y}(1)^2 - \sum_i Y_i(1)^2}{N_t^2} + \frac{N^2\bar{Y}(0)\bar{Y}(1) - \sum_i Y_i(0)Y_i(1)}{N_t N_c} + \frac{N^2\bar{Y}(1)\bar{Y}(0) - \sum_i Y_i(0)Y_i(1)}{N_t N_c} + \frac{N^2\bar{Y}(0)^2 - \sum_i Y_i(0)^2}{N_c^2}\right)\right]$$

$$= \frac{1}{N^2}\left[\frac{N_c}{N_t}\left(1 + \frac{1}{N-1}\right)\sum_i Y_i(1)^2 - \frac{N^2}{N-1}\frac{N_c}{N_t}\bar{Y}(1)^2 + \frac{N_t}{N_c}\left(1 + \frac{1}{N-1}\right)\sum_i Y_i(0)^2 - \frac{N^2}{N-1}\frac{N_t}{N_c}\bar{Y}(0)^2 + 2\sum_i Y_i(0)Y_i(1)\left(1 + \frac{1}{N-1}\right) - \frac{N^2}{N-1}2\cdot\bar{Y}(0)\bar{Y}(1)\right]$$

$$= \frac{N_c + N_t - N_t}{N_t}\cdot\frac{1}{N-1}\cdot\frac{1}{N}\left(\sum_i Y_i(1)^2 - N\bar{Y}(1)^2\right) + \frac{N_t + N_c - N_c}{N_c}\cdot\frac{1}{N-1}\cdot\frac{1}{N}\left(\sum_i Y_i(0)^2 - N\bar{Y}(0)^2\right) + \frac{1}{N-1}\cdot\frac{1}{N}\left(2\sum_i Y_i(0)Y_i(1) - 2N\bar{Y}(0)\bar{Y}(1)\right).$$

$$= \frac{S_t^2}{N_t} + \frac{S_c^2}{N_c} - \frac{S_t^2}{N} - \frac{S_c^2}{N} + \frac{1}{N-1}\cdot\frac{1}{N}\left(2\sum_i Y_i(0)Y_i(1) - 2N\bar{Y}(0)\bar{Y}(1)\right)$$

$$= \frac{S_t^2}{N_t} + \frac{S_c^2}{N_c} - \frac{1}{N}\cdot\frac{1}{N-1}\left(\sum_i Y_i(1)^2 + \sum_i Y_i(0)^2 - N\bar{Y}(1)^2 - N\bar{Y}(0)^2 + 2N\bar{Y}(0)\bar{Y}(1) - 2\sum_i Y_i(0)Y_i(1)\right)$$

$$= \frac{S_t^2}{N_t} + \frac{S_c^2}{N_c} - \frac{S_{tc}^2}{N}$$

Done!

$$Y_i^+ = \frac{N}{N_t}Y_i(1) + \frac{N}{N_c}Y_i(0)$$

$$E_w(D_i^2) = \frac{N_c N_t}{N^2}$$

$$E_w(D_i D_j) = -\frac{N_c N_t}{N^2(N-1)}$$

$$N\bar{Y}(1) = \sum_i Y_i(1)$$

$$\sum_{j\neq i} Y_j(1) = N\bar{Y}(1) - Y_i(1)$$

$$S_t^2 = \frac{1}{N-1}\left(\sum_i Y_i(1)^2 - N\bar{Y}(1)^2\right)$$

The Sampling Variance (II): Alternative Representation

$$\mathbb{V}_W\left(\overline{Y}_{\mathrm{t}}^{\mathrm{obs}}-\overline{Y}_{\mathrm{c}}^{\mathrm{obs}}\right)=\frac{N_{\mathrm{t}}}{N\cdot N_{\mathrm{c}}}\cdot S_c^2+\frac{N_{\mathrm{c}}}{N\cdot N_{\mathrm{t}}}\cdot S_t^2+\frac{2}{N}\cdot\rho_{tc}\cdot S_c\cdot S_t$$

$$S_c^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(0)-\overline{Y}(0)\right)^2,\quad S_t^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)^2,$$

$$\rho_{tc}=\frac{1}{(N-1)\cdot S_c\cdot S_t}\sum_{i=1}^{N}\left(Y_i(1)-\overline{Y}(1)\right)\cdot\left(Y_i(0))-\overline{Y}(0)\right)$$

ρ_{tc} → the correlation coefficient

Proof for (II)

$$\mathrm{Var}(\hat{\tau}^{dif}) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i^{+}\cdot D_i\right) = \frac{1}{N^2}\cdot E\left[\left(\sum_{i=1}^{N} Y_i^{+}\cdot D_i\right)^2\right]$$

$$= \frac{1}{N^2}\left[\sum_i (Y_i^{+})^2 \cdot ED_i^2 + \sum_i\sum_{j\neq i} Y_i^{+}Y_j^{+}\cdot E(D_iD_j)\right]$$

$$= \frac{1}{N^2}\left[\sum_i (Y_i^{+})^2\cdot\frac{N_tN_c}{N^2} - \sum_i\sum_{j\neq i} Y_i^{+}Y_j^{+}\cdot\frac{N_tN_c}{N^2(N-1)}\right] = \frac{N_tN_c}{N^4}\left[\sum_i (Y_i^{+})^2 - \sum_i\sum_{j\neq i} Y_i^{+}Y_j^{+}\cdot\frac{1}{N-1}\right]$$

$$= \frac{N_tN_c}{N^4}\cdot\frac{1}{N-1}\left\{(N-1)\sum_i\left(\frac{N}{N_t}Y_i(1)+\frac{N}{N_c}Y_i(0)\right)^2 - \sum_i\sum_{j\neq i}\left(\frac{N}{N_t}Y_i(1)+\frac{N}{N_c}Y_i(0)\right)\cdot\left(\frac{N}{N_t}Y_j(1)+\frac{N}{N_c}Y_j(0)\right)\right\}$$

Aside: ① $\because \sum_i (x_i-\bar{x})^2 = \frac{1}{n}\left[(n-1)\sum x_i^2 - \sum_i\sum_{j\neq i} x_ix_j\right]$

$\therefore \sum_i (Y_i(1)-\bar{Y}(1))^2 = \frac{1}{N}\left[(N-1)\sum_i Y_i^2(1) - \sum_i\sum_{j\neq i} Y_i(1)\cdot Y_j(1)\right]$

↘ 分成 ① 只有 $Y_i(1)$ ② 只有 $Y_i(0)$ ③ 只有 $Y_i(0)\cdot Y_k(1)$ 交叉项 三部分.

② 同上理由, $\sum_i (Y_i(0)-\bar{Y}(0))^2 = \frac{1}{N}\left[(N-1)\sum_i Y_i^2(0) - \sum_i\sum_{j\neq i} Y_i(0)\cdot Y_j(0)\right]$

③ $\because \sum_i (x_i-\bar{x})(y_i-\bar{y}) = \sum_i (x_iy_i - n\bar{x}\bar{y}) = \frac{1}{n}\left[(n-1)\sum_i x_iy_i - \sum_i\sum_{j\neq i} x_iy_j\right]$

$\therefore \sum_i (Y_i(1)-\bar{Y}(1))(Y_i(0)-\bar{Y}(0)) = \frac{1}{N}\left[(N-1)\sum_i Y_i(1)Y_i(0) - \sum_i\sum_{j\neq i}(Y_i(0)\cdot Y_j(1))\right]$

$$= \frac{1}{N}\left[(N-1)\sum_i Y_i(1)Y_i(0) - \sum_i\sum_{j\neq i}(Y_j(0)\cdot Y_i(1))\right]$$

$$= \frac{N_tN_c}{N^4}\cdot\frac{1}{N-1}\cdot\left(\frac{N}{N_t}\right)^2\cdot N\cdot\sum_i (Y_i(1)-\bar{Y}(1))^2 + \frac{N_tN_c}{N^4}\cdot\frac{1}{N-1}\cdot\left(\frac{N}{N_c}\right)^2\cdot N\cdot\sum_i (Y_i(0)-\bar{Y}(0))^2$$

$$+ \frac{N_tN_c}{N^4}\cdot\frac{1}{N-1}\cdot\frac{N^2}{N_tN_c}\cdot\left\{(N-1)\cdot 2\sum_i Y_i(1)Y_i(0) - \sum_i\sum_{j\neq i}\left[Y_i(1)\cdot Y_j(0) + Y_i(0)Y_j(1)\right]\right\}$$

$$= \frac{N_c}{N\cdot N_t}\cdot\underbrace{\frac{1}{N-1}\sum_i (Y_i(1)-\bar{Y}(1))^2}_{S_t^2} + \frac{N_t}{N\cdot N_c}\cdot\underbrace{\frac{1}{N-1}\sum_i (Y_i(0)-\bar{Y}(0))^2}_{S_c^2} + \frac{1}{N}\cdot\frac{1}{N-1}\cdot\underbrace{2\sum_i (Y_i(1)-\bar{Y}(1))(Y_i(0)-\bar{Y}(0))}_{\rho_{tc}\cdot S_c\cdot S_t}$$

$$= \frac{N_c}{N\cdot N_t}\cdot S_t^2 + \frac{N_t}{N\cdot N_c}\cdot S_c^2 + \frac{2}{N}\cdot\rho_{tc}\cdot S_c\cdot S_t$$