Module 5 : Neymanian inference

Marie-Abèle Bind

STAT 140

Spring 2026

Neyman's definition of confidence interval (CI) (1934, page 590)

Suppose we are taking samples, Σ, from some population π. We are interested in a certain collective character of this population, say θ. Denote by x a collective character of the sample Σ and suppose that we have been able to deduce its frequency distribution, say $p(x|\theta)$, in repeated samples and that this is dependent on the unknown collective character, θ, of the population π....

Denote now by $\varphi(\theta)$ the unknown probability distribution *a priori* of θ...

...[T]he probability of our being wrong is less than or at most equal to $1 - \varepsilon$, and this whatever the probability law *a priori*, $\varphi(\theta)$.

The value of ε, chosen in a quite arbitrary manner, I propose to call the "confidence coefficient." If we choose, for instance, $\varepsilon = \cdot 99$ and find for every possible x the intervals $[\theta_1(x), \theta_2(x)]$ having the properties defined, we could roughly describe the position by saying that we have 99 per cent. confidence in the fact that θ is contained between $\theta_1(x)$ and $\theta_2(x)$....

... [I] call the intervals $[\theta_1(x), \theta_2(x)]$ the confidence intervals, corresponding to the confidence coefficient ε.

Jerzey Neyman

- Founding father of the potential outcomes in randomized experiments.
- Focused on methods for the estimation of, and inference for, average causal effects (ACE) :

$$\tau = \frac{1}{N}\sum_{i}^{N}\Big(Y_i(W_i = 1) - Y_i(W_i = 0)\Big) = \overline{Y(1)} - \overline{Y(0)}$$

- Used randomization distribution possibly in combination with random sampling of the units in the experiment from a larger population of units.
- Interested in the long-run operating characteristics of statistical procedures.
- Estimation and inference : point estimator and interval estimator with good properties in large samples and in the long run.

Confidence Intervals

- (100-α)% Confidence Intervals (CI) for τ :

$$CI_{(100-\alpha)\%} = \left[\hat{\tau} - t_{\frac{\alpha}{2},df} \sqrt{\widehat{Var}(\hat{\tau})} \; ; \; \hat{\tau} + t_{1-\frac{\alpha}{2},df} \sqrt{\widehat{Var}(\hat{\tau})} \right]$$

Data from the Perfect Doctor (Rubin's example)

- Let us look at this example to get the intuition of what Neyman was after.

Data from the Perfect Doctor (Rubin's example)

- Let us look at this example to get the intuition of what Neyman was after.

- W_i : surgery ($W_i = 0$ if unit i receives standard surgery, $W_i = 1$ if unit i receives new surgery)

Data from the Perfect Doctor (Rubin's example)

- Let us look at this example to get the intuition of what Neyman was after.
- W_i : surgery ($W_i = 0$ if unit i receives standard surgery, $W_i = 1$ if unit i receives new surgery)
- Y_i : years that unit i lived after receiving surgery W_i

Science Table

Assume we know the Science table (i.e., the "Truth") :

unit i	$Y_i(W_i\!=\!0)$	$Y_i(W_i\!=\!1)$
1	13	14
2	6	0
3	4	1
4	5	2
5	6	3
6	6	1
7	8	10
8	8	9

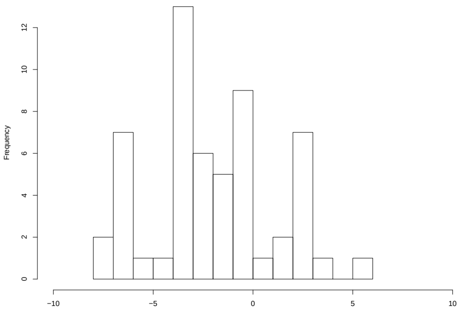
The true average causal effect $(\tau) := \overline{Y(1)} - \overline{Y(0)} = -2$.

All possible allocations

W	$\hat{\tau}(W)$
(1,1,1,0,0,0,0,0)	-1.6
...	...
...	...
...	...
...	...
...	...
...	...
...	...
...	...
...	...
...	...
...	...
...	...
(0,0,0,0,0,1,1,1)	-0.1

w	$\overline{y_1} - \overline{y_0}$
11100000	-1.6
11010000	-1.1
11001000	-0.5
11000100	-1.2
11000010	2.2
11000001	1.9
10110000	-1.1
10101000	-0.6
10100100	-1.3
10100010	2.1
10100001	1.8
10011000	-0.1
10010100	-0.7
10010010	2.7
10010001	2.3
10001100	-0.2
10001010	3.2
10001001	2.9
10000110	2.5
10000101	2.2
10000011	5.6
01110000	-7.2
01101000	-6.7
01100100	-7.3
01100010	-3.9
01100001	-4.3
01011000	-6.1
01010100	-6.8
01010010	-3.4
01010001	-3.7
01001100	-6.3
01001010	-2.9
01001001	-3.2
01000110	-3.5
01000101	-3.9
01000011	-0.5
00111000	-6.2
00110100	-6.9
00110010	-3.5
00110001	-3.8
00101100	-6.3
00101010	-2.9
00101001	-3.3
00100110	-3.6
00100101	-3.9
00100011	-0.5
00011100	-5.8
00011010	-2.4
00011001	-2.7
00010110	-3.1
00010101	-3.4
00010011	0.0
00001110	-2.5
00001101	-2.9
00001011	0.5
00000111	-0.1

Distribution of $\hat{\tau}(W)$



- Mean $:= \frac{1}{56}\sum\hat{\tau}(W) = -2$; Median = -2.63
- 2.5^{th} quantile= -7.075 and 97.5^{th} quantile= 3.075
- Variance=9.75 and standard deviation=3.12

Conduct a single randomized experiment

A completely randomized experiment is conducted for which this W^{obs}=(1,0,0,0,0,0,1,1) was observed :

W_i^{obs}	$Y_i(W_i=0)$	$Y_i(W_i=1)$
1	?	14
0	6	?
0	4	?
0	5	?
0	6	?
0	6	?
1	?	10
1	?	9

The estimated average causal effect (i.e., $\hat{\tau}(W^{obs})) := \overline{Y}_1^{obs}$-$\overline{Y}_0^{obs}= 5.6$.

95% Confidence Interval (CI) for τ : [-0.46 ;11.66]

An estimator of τ

- Suppose that we observe data from a completely randomized experiment in which $\mathrm{N}_1 = \sum_{i=1}^{N} = W_i$ units are randomly selected to be assigned to treatment and the remaining $\mathrm{N}_0 = \sum_{i=1}^{N} = (1 - W_i)$ to control.

An estimator of τ

- A natural estimator for the ACE is :

$$\hat{\tau} = \overline{Y_1^{obs}} - \overline{Y_0^{obs}}$$

  - $\overline{Y}_1^{obs} = \frac{1}{N_1} \sum_{i:W_i=1} Y_i^{obs}$
  - $\overline{Y}_0^{obs} = \frac{1}{N_0} \sum_{i:W_i=0} Y_i^{obs}$

Sampling variance of $\hat{\tau}$

- $\text{Var}(\hat{\tau}) = \ldots = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{0,1}^2}{N}$

Sampling variance of $\hat{\tau}$

- $\text{Var}(\hat{\tau}) = ... = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{0,1}^2}{N}$
  - $S_0^2 = \frac{1}{N-1} \sum_{i=1}^{N} \Big(Y_i(W_i = 0) - \overline{Y}(0) \Big)^2$

Sampling variance of $\hat{\tau}$

- $\text{Var}(\hat{\tau}) = ... = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{0,1}^2}{N}$

  - $S_1^2 = \frac{1}{N-1} \sum_{i=1}^{N} \Big(Y_i(W_i = 1) - \overline{Y}(1) \Big)^2$

Sampling variance of $\hat{\tau}$

- $\text{Var}(\hat{\tau}) = \ldots = \frac{S_0^2}{N_0} + \frac{S_1^2}{N_1} - \frac{S_{0,1}^2}{N}$

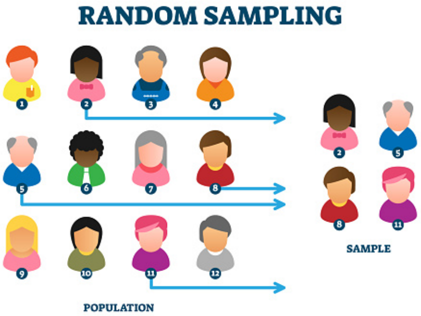
- $S_{0,1}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(W_i=1) - Y_i(W_i=0) - \left(\overline{Y}(1) - \overline{Y}(0)\right)\right)^2$
  $= \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i(W_i=1) - Y_i(W_i=0) - \tau\right)^2$

An estimator of the sampling variance of the ACE

- Assume that for each unit, $Y_i(W_i = 1) - Y_i(W_i = 0)$ is additive and constant :
$$\widehat{Var}(\hat{\tau}) = \frac{s_0^2}{N_0} + \frac{s_1^2}{N_1}$$

- Widely used, even when the assumption of an additive and constant effect may be known to be inaccurate.

Random sampling



Random sampling

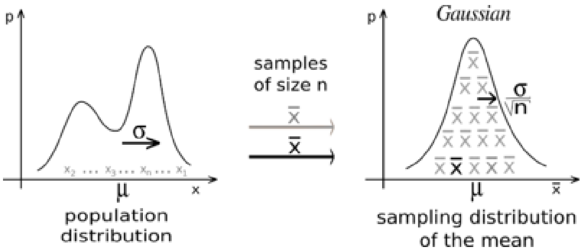
- Suppose we have a population of size N from which we want to learn about the effect of a treatment W on an outcome Y.

unit i	$Y_i(W_i{=}0)$	$Y_i(W_i{=}1)$
1	$Y_1(W_1{=}0)$	$Y_1(W_1{=}1)$
...	...	...
...	...	...
...	...	...
...	...	...
...	...	...
...	...	...
N	$Y_N(W_N{=}0)$	$Y_N(W_N{=}1)$

- Suppose we conduct a study in a random sample of size n of the population (typically, $n << N$).

unit i	W_i^{obs}	$Y_i(W_i{=}0)$	$Y_i(W_i{=}1)$	Y_i^{obs}
1	W_1^{obs}	$Y_1(W_1{=}0)$	$Y_1(W_1{=}1)$	Y_1^{obs}
...	...	...	...	
n	W_n^{obs}	$Y_n(W_n{=}0)$	$Y_n(W_n{=}1)$	Y_n^{obs}
n+1		$Y_{n+1}(W_{n+1}{=}0)$	$Y_{n+1}(W_{n+1}{=}1)$	
...				
...				
...				
N		$Y_N(W_N{=}0)$	$Y_N(W_N{=}1)$	

Central Limit Theorem (CLT)



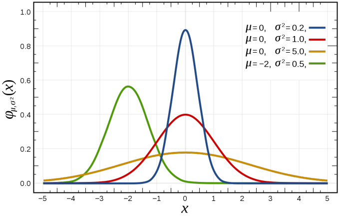
Central Limit Theorem (CLT)

- In large sample, $\frac{\hat{\tau}-\tau}{\sqrt{\widehat{Var}(\hat{\tau})}}$ can be approximated by known distributions.

Central Limit Theorem (CLT)

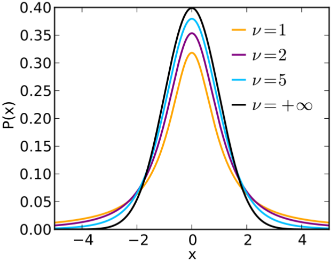
- In large sample, $\frac{\hat{\tau}-\tau}{\sqrt{\widehat{Var}(\hat{\tau})}}$ can be approximated by known distributions.

- Normal distribution, which is specified with a mean (here, μ) and variance (here, σ^2) :



Central Limit Theorem (CLT)

- In large sample, $\frac{\hat{\tau}-\tau}{\sqrt{\widehat{Var}(\hat{\tau})}}$ can be approximated by known distributions.
- Student's t distribution, which is specified using degrees of freedom (here, ν) :



Confidence Intervals

- Neyman is interested in estimating the precision of the estimator for the average causal effect τ.

Confidence Intervals

- Neyman is interested in estimating the precision of the estimator for the average causal effect τ.
- A confidence interval with confidence coefficient $1-\alpha$ is a pair of function $C_L(Y^{obs}, W)$ and $C_U(Y^{obs}, W)$, defining an interval $[C_L(Y^{obs}, W)\,;\, C_U(Y^{obs}, W)]$ such that :

$$P(C_L(Y^{obs}, W) \leq \tau \leq C_U(Y^{obs}, W)) \geq 1-\alpha.$$

Confidence Intervals

- (100-α)% Confidence Interval (CI) for τ :

$$CI_{(100-\alpha)\%} = \left[\; \hat{\tau} + t_{\frac{\alpha}{2},df} \; \sqrt{\widehat{Var}(\hat{\tau})} \; ; \; \hat{\tau} + t_{1-\frac{\alpha}{2},df} \; \sqrt{\widehat{Var}(\hat{\tau})} \; \right]$$

Confidence Intervals

- 95% Confidence Intervals (CI) for τ :

$$CI_{95\%} = \left[\; \hat{\tau} + t_{0.025,df} \; \sqrt{\widehat{Var}(\hat{\tau})} \; ; \; \hat{\tau} + t_{0.975,df} \; \sqrt{\widehat{Var}(\hat{\tau})} \; \right]$$

Confidence Intervals

- 95% Confidence Intervals (CI) for τ :

$$CI_{95\%} = \left[\hat{\tau} + t_{0.025,df} \sqrt{\widehat{Var}(\hat{\tau})} \; ; \; \hat{\tau} + t_{0.975,df} \sqrt{\widehat{Var}(\hat{\tau})} \right]$$

$$df = \frac{\left(\frac{s_0^2}{N_0} + \frac{s_1^2}{N_1}\right)^2}{\frac{\left(\frac{s_0^2}{N_0}\right)^2}{N_0 - 1} + \frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1}}$$

Confidence Intervals

- 95% Confidence Intervals (CI) for τ :

$$CI_{95\%} = \left[\ \hat{\tau} + t_{0.025,df} \ \sqrt{\widehat{Var}(\hat{\tau})} \ ; \ \hat{\tau} + t_{0.975,df} \ \sqrt{\widehat{Var}(\hat{\tau})} \ \right]$$

- 95% Confidence Intervals (CI) for τ (assuming a fairly large sample) :

$$CI_{95\%} = \left[\ \hat{\tau} - 1.96 \ \sqrt{\widehat{Var}(\hat{\tau})} \ ; \ \hat{\tau} + 1.96 \ \sqrt{\widehat{Var}(\hat{\tau})} \ \right]$$

Neyman's definition of CI (1934, page 590)

Suppose we are taking samples, Σ, from some population π. We are interested in a certain collective character of this population, say θ. Denote by x a collective character of the sample Σ and suppose that we have been able to deduce its frequency distribution, say $p(x|\theta)$, in repeated samples and that this is dependent on the unknown collective character, θ, of the population π....

Denote now by $\varphi(\theta)$ the unknown probability distribution *a priori* of θ...

...[T]he probability of our being wrong is less than or at most equal to $1 - \varepsilon$, and this whatever the probability law *a priori*, $\varphi(\theta)$.

The value of ε, chosen in a quite arbitrary manner, I propose to call the "confidence coefficient." If we choose, for instance, $\varepsilon = \cdot 99$ and find for every possible x the intervals $[\theta_1(x), \theta_2(x)]$ having the properties defined, we could roughly describe the position by saying that we have 99 per cent. confidence in the fact that θ is contained between $\theta_1(x)$ and $\theta_2(x)$....

... [I] call the intervals $[\theta_1(x), \theta_2(x)]$ the confidence intervals, corresponding to the confidence coefficient ε.

Hypothesis testing

- Neyman used the sampling variance estimate to carry out test of hypotheses concerning the average treatment effect τ.
- Neyman's null hypothesis :

$$\frac{1}{N}\sum_{i=1}^{N}[Y_i(W_i = 1) - Y_i(W_i = 0)] = 0$$

- One possible test statistic is : $\frac{\hat{\tau}}{\sqrt{\widehat{Var}(\hat{\tau})}}$, whose distribution under the Neyman's null hypothesis in fairly large sample can be approximated by a Student's t distribution (with df described 3 slides ago).
- An approximating p-value can be calculated using the Student's t distribution (with df described 3 slides ago).