

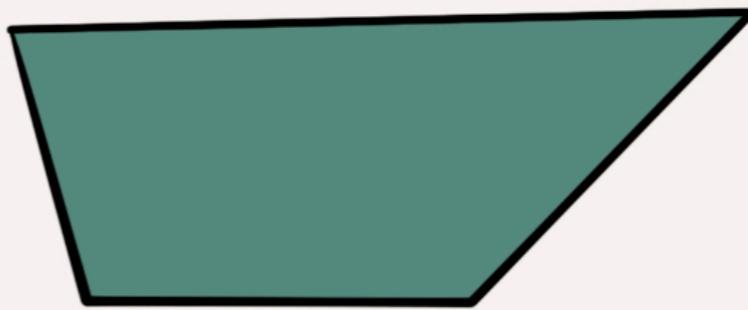
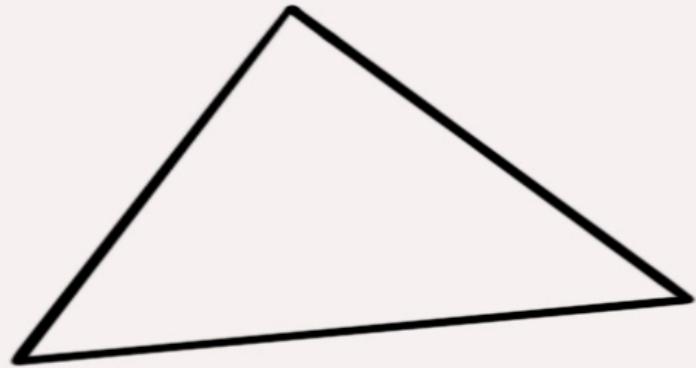
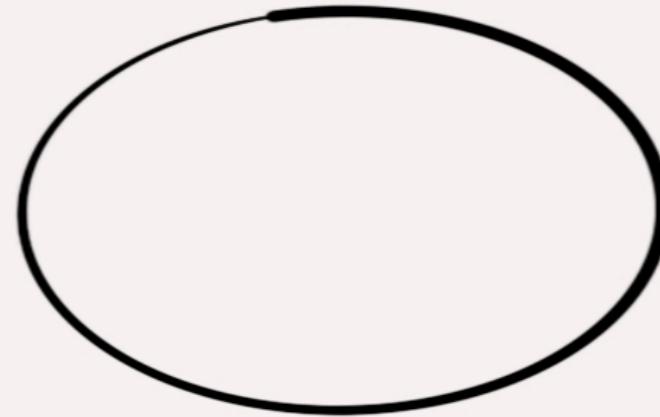
Statistics & Causal Inference:  
Neyman { Fisher



Jake Bowers  
<http://jakebowers.org>

EGAP 2021  
19 July 2021

What is the causal effect?



What is the causal effect?



$$A \quad Y_A = 21 \quad T_A = 1$$

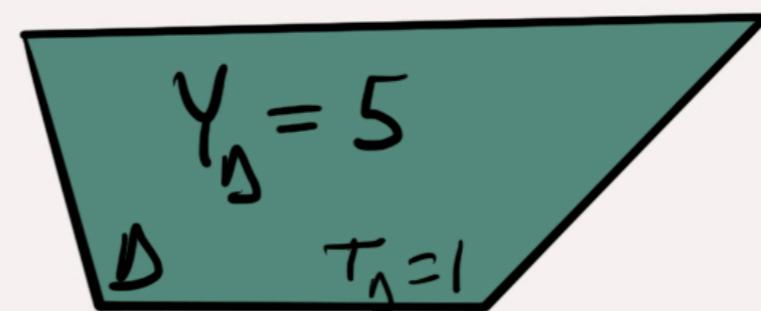
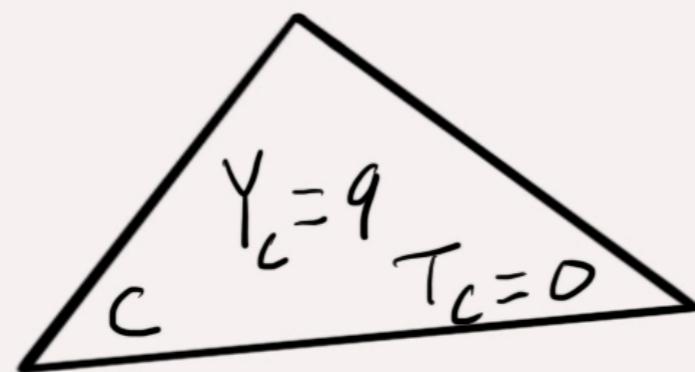
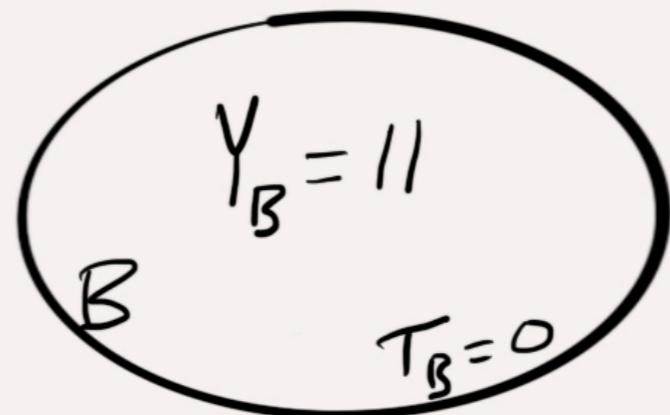
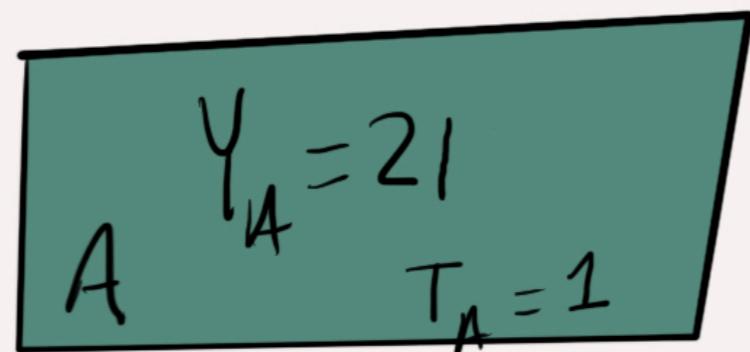
$$B \quad Y_B = 11 \quad T_B = 0$$

$$C \quad Y_C = 9 \quad T_C = 0$$

$$D \quad Y_D = 5 \quad T_D = 1$$

$i$ Unit	$T$ Treatment	$Y$ Observed Outcome
A	1	21
B	0	11
C	0	9
D	1	5

What is the causal effect?



$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect( $\tau$ )	We don't know!
A	1	21	21	?	$21 - ? = \tau_A$	
B	0	11	?	11	$? - 11 = \tau_B$	
C	0	9	?	9	$? - 9 = \tau_C$	
D	1	5	5	?	$5 - ? = \tau_D$	



$$A \quad Y_A = 21 \quad T_A = 1$$

$$B \quad Y_B = 11 \quad T_B = 0$$

$$C \quad Y_C = 9 \quad T_C = 0$$



I don't know about the individual causal effects. BUT  
I can help you ESTIMATE the AVERAGE causal effect!

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect( $\gamma$ )
A	1	21	21	?	$21 - ? = \gamma_A$
B	0	11	?	11	$? - 11 = \gamma_B$
C	0	9	?	9	$? - 9 = \gamma_C$
D	1	5	5	?	$5 - ? = \gamma_D$

Define:  $\bar{\gamma} = (\gamma_A + \gamma_B + \gamma_C + \gamma_D) / 4$  (The avg. causal effect)

Notice:  $\gamma_A = Y_A(1) - Y_A(0)$ ,  $\gamma_B = Y_B(1) - Y_B(0)$ , ...

$$\begin{aligned}\bar{\gamma} &= ((Y_A(1) - Y_A(0)) + \dots + (Y_D(1) - Y_D(0))) / 4 \\ &= (Y_A(1) + \dots + Y_D(1)) / 4 - (Y_A(0) + \dots + Y_D(0)) / 4\end{aligned}$$

$$A \quad Y_A = 21 \quad T_A = 1$$

$$B \quad Y_B = 11 \quad T_B = 0$$

$$C \quad Y_C = 9 \quad T_C = 0$$



I don't know about the individual causal effects. BUT  
I can help you ESTIMATE the AVERAGE causal effect!

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect( $\gamma$ )
A	1	21	21	?	$21 - ? = \gamma_A$
B	0	11	?	11	$? - 11 = \gamma_B$
C	0	9	?	9	$? - 9 = \gamma_C$
D	1	5	5	?	$5 - ? = \gamma_D$

Define:  $\bar{\gamma} = (\gamma_A + \gamma_B + \gamma_C + \gamma_D)/4$  (The avg. causal effect)  
 an Unobserved Estimand  $= (Y_A(1) + \dots + Y_D(1))/4 - (Y_A(0) + \dots + Y_D(0))/4$

$$\bar{\gamma} = \bar{Y}(1) - \bar{Y}(0)$$

$$A \quad Y_A = 21 \quad T_A = 1$$

$$B \quad Y_B = 11 \quad T_B = 0$$

$$C \quad Y_C = 9 \quad T_C = 0$$



I don't know about the individual causal effects. BUT  
I can help you ESTIMATE the AVERAGE causal effect!

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect( $\gamma$ )
A	1	21	21	?	$21 - ? = \gamma_A$
B	0	11	?	11	$? - 11 = \gamma_B$
C	0	9	?	9	$? - 9 = \gamma_C$
D	1	5	5	?	$5 - ? = \gamma_D$

Define

an

Unobserved

$$\bar{\gamma} = (\gamma_A + \gamma_B + \gamma_C + \gamma_D) / 4$$

Estimand

$$\bar{\gamma} = \bar{Y}(1) - \bar{Y}(0)$$

choose

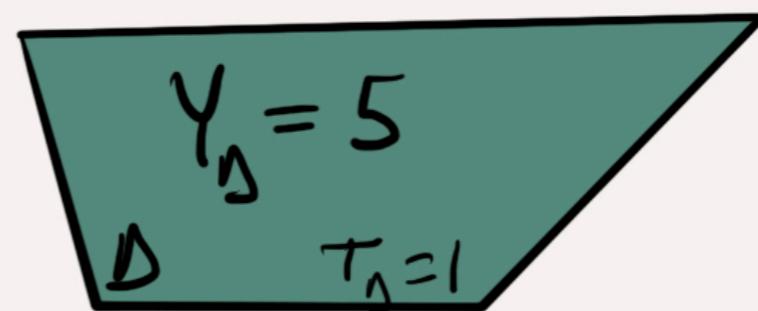
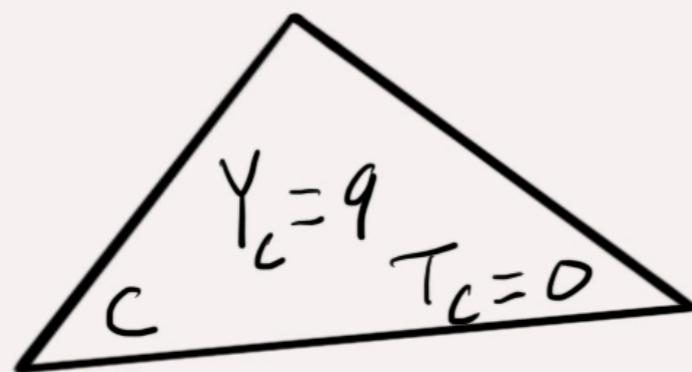
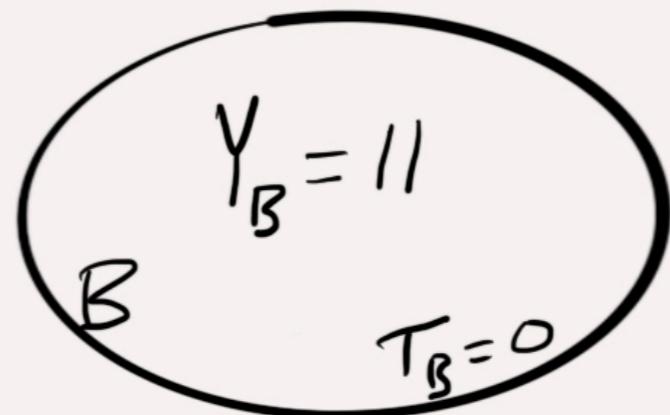
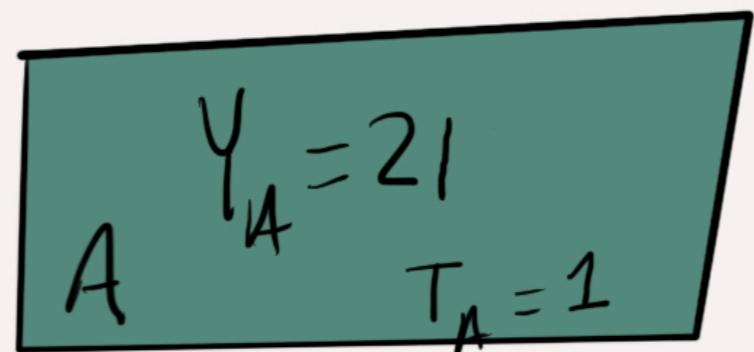
an

Estimator

$$\hat{\bar{\gamma}} = \hat{Y}(1) - \hat{Y}(0) = \frac{(21+5)}{2} - \frac{(11+9)}{2} = 13 - 10 = 3$$

(Show (Later) that  $\hat{\bar{\gamma}}$  is a good estimator of  $\bar{\gamma}$ )

What is the causal effect?



$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect( $\tau$ )	We don't know!
A	1	21	21	?	$21 - ? = \tau_A$	
B	0	11	?	11	$? - 11 = \tau_B$	
C	0	9	?	9	$? - 9 = \tau_C$	
D	1	5	5	?	$5 - ? = \tau_D$	



$$A \quad Y_A = 21 \quad T_A = 1$$

$$B \quad Y_B = 11 \quad T_B = 0$$

$$C \quad Y_C = 9 \quad T_C = 0$$

$$D \quad Y_D = 5 \quad T_D = 1$$



I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

An idea: "No effects"  
"H<sub>0</sub>: Y<sub>i</sub>(1) = Y<sub>i</sub>(0)"

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect ( $\tau$ )
A	1	21	21	?	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	?	$5 - ? = \tau_D$

A  $Y_A = 21$   
 $T_A = 1$

B  $Y_B = 11$   
 $T_B = 0$

C  $Y_C = 9$   
 $T_C = 0$

D  $Y_D = 5$   
 $T_D = 1$

—o—  
|m|

I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

An idea: "No effects"  
("H<sub>0</sub>:  $Y_i(1) = Y_i(0)$ ")

What does this idea imply?

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect ( $\tau$ )
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	5	$5 - ? = \tau_D$

Notice:  $Y_i = T_i Y_i(1) + (1-T_i) Y_i(0)$

So  $H_0$  implies:  $Y_i = T_i \{Y_i(0)\} + (1-T_i) Y_i(0)$   
 $= Y_i(0)$

A  $Y_A = 21$   
 $T_A = 1$

B  $Y_B = 11$   
 $T_B = 0$

C  $Y_C = 9$   
 $T_C = 0$

D  $Y_D = 5$   
 $T_D = 1$



I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

An idea: "No effects"  
("H<sub>0</sub>:  $Y_i(1) = Y_i(0)$ ")  
 $H_0 \Rightarrow Y_i = Y_i(0)$   
observed

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect ( $\tau$ )
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	5	$5 - ? = \tau_D$

So:  $\frac{(21+5)}{2} - \frac{(11+9)}{2} = 3$  is compatible with  $H_0$  even if not  $\emptyset$ .

But, A  $\setminus$  D are only treated by chance.

If A  $\setminus$  B were treated,  $H_0 \Rightarrow \frac{(21+11)}{2} - \frac{(9+5)}{2} = 9$

If C  $\setminus$  D were treated,  $H_0 \Rightarrow \frac{(9+5)}{2} - \frac{(21+11)}{2} = -9$

$$A \quad Y_A = 21 \quad T_A = 1$$

$$B \quad Y_B = 11 \quad T_B = 0$$

$$C \quad Y_C = 9 \quad T_C = 0$$

$$D \quad Y_D = 5 \quad T_D = 1$$



I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

① Idea /  $H_0: Y_i(1) = Y_i(0)$

②  $H_0$  implies  $Y_i = Y_i(0)$

③  $H_0$  implies that

a test statistic

summarizing  $T \rightarrow \{Y\}$

$$t(T, Y) = \left( \frac{\sum 1+5}{2} \right) - \left( \frac{11+9}{2} \right) \\ = 3$$

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect ( $\gamma$ )
A	1	21	21	21	$21 - ? = \gamma_A$
B	0	11	?	11	$? - 11 = \gamma_B$
C	0	9	?	9	$? - 9 = \gamma_C$
D	1	5	5	5	$5 - ? = \gamma_D$

But also these other possible  $t(T, Y)$ s:

$i$	$\tilde{T}_1$	$\tilde{T}_2$	$\tilde{T}_3$	$\tilde{T}_4$	$\tilde{T}_5$	$\tilde{T}_6$
A	1	1	1	0	0	0
B	0	1	0	1	1	0
C	0	0	1	1	0	1
D	1	0	0	0	1	1

→ 3 9 7 -3 -7 -9

A  $Y_A = 21$   
 $T_A = 1$

B  $Y_B = 11$   
 $T_B = 0$

C  $Y_C = 9$   
 $T_C = 0$

D  $Y_D = 5$   
 $T_D = 1$



I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

① Idea/H<sub>0</sub>:  $Y_i(1) = Y_i(0)$

② H<sub>0</sub> implies  $Y_i = Y_i(0)$

③ H<sub>0</sub> and Randomization

imply  $\leqq$  equally

likely  $t(T, Y)$

values.

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect ( $\tau$ )
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	5	$5 - ? = \tau_D$

In this experiment,  
"no effects"  
means

$i$	$\tilde{T}_1$	$\tilde{T}_2$	$\tilde{T}_3$	$\tilde{T}_4$	$\tilde{T}_5$	$\tilde{T}_6$
A	1	1	1	0	0	0
B	0	1	0	1	1	0
C	0	0	1	1	0	1
D	1	0	0	0	1	1

D 3 9 7 -3 -7 -9

$t(T, Y)$   
no effects

A  $Y_A = 21$   
 $T_A = 1$

B  $Y_B = 11$   
 $T_B = 0$

C  $Y_C = 9$   
 $T_C = 0$

D  $Y_D = 5$   
 $T_D = 1$



I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

① Idea/H<sub>0</sub>:  $Y_i(1) = Y_i(0)$

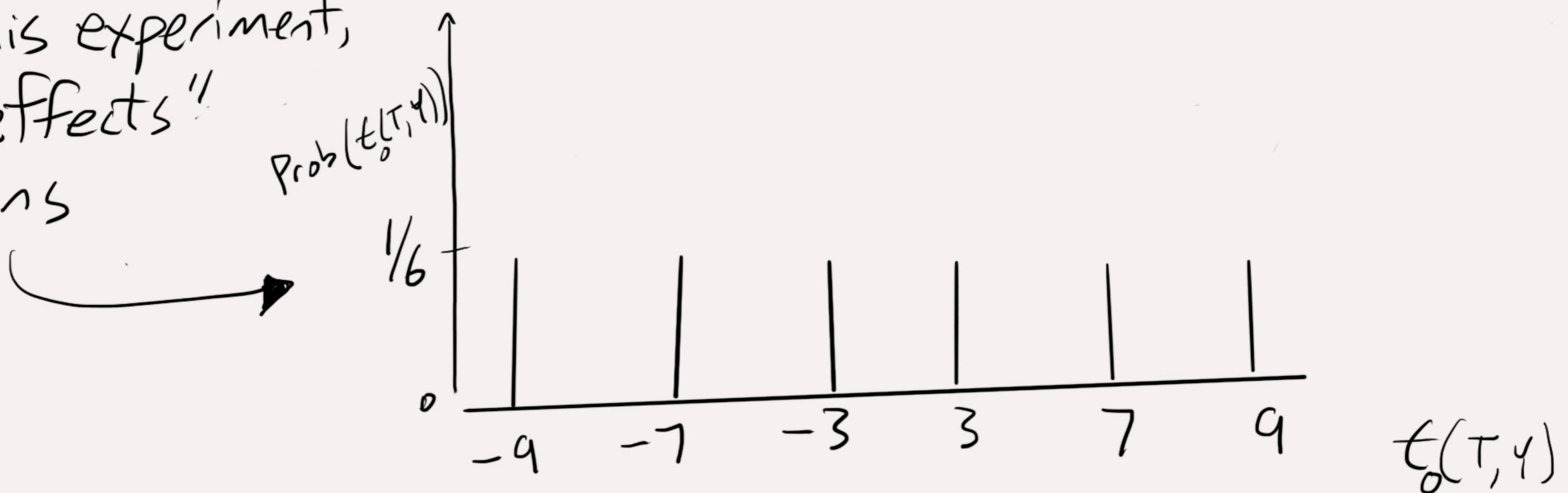
② H<sub>0</sub> implies  $Y_i = Y_i(0)$

③ H<sub>0</sub> and Randomization

imply  $\leqq$  equally likely  $t_0(T, Y)$  values.

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect ( $\tau$ )
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	5	$5 - ? = \tau_D$

In this experiment,  
"no effects"  
means



A  $Y_A = 21$   
 $T_A = 1$

B  $Y_B = 11$   
 $T_B = 0$

C  $Y_C = 9$   
 $T_C = 0$

D  $Y_D = 5$   
 $T_D = 1$



I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

① Idea/H<sub>0</sub>:  $Y_i(1) = Y_i(0)$

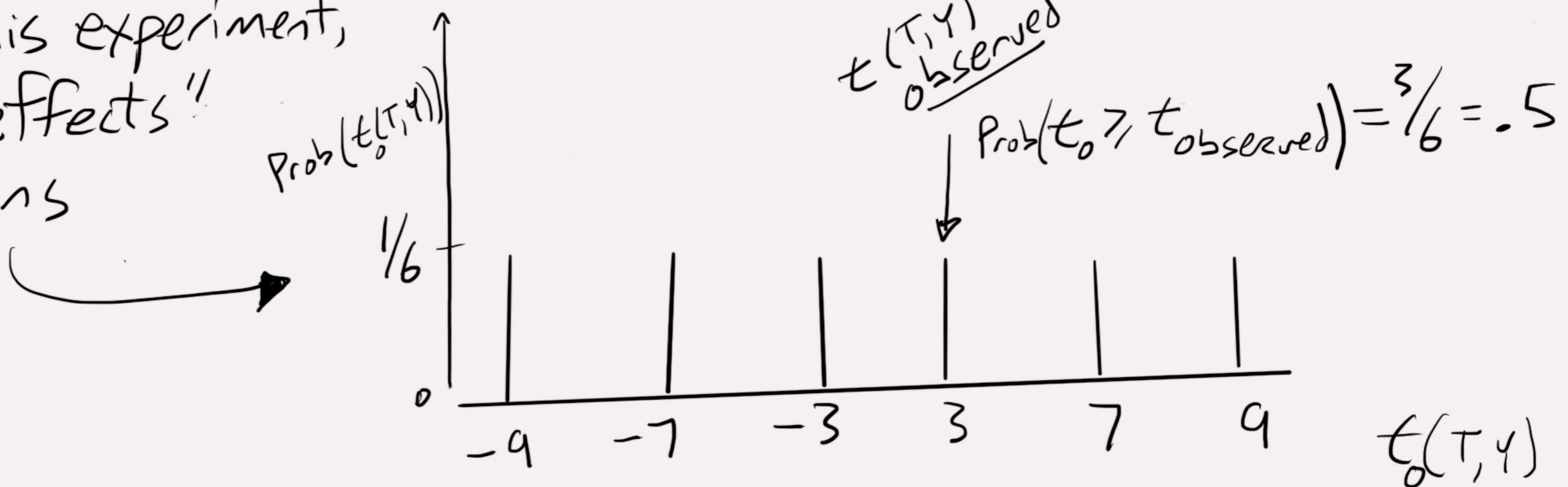
② H<sub>0</sub> implies  $Y_i = Y_i(0)$

③ H<sub>0</sub> and Randomization

imply  $\leqq$  equally likely  $t_0(T, Y)$  values.

$i$	$T$	$Y$	$Y(T=1)$	$Y(T=0)$	Effect ( $\tau$ )
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	5	$5 - ? = \tau_D$

In this experiment,  
"no effects"  
means



A  $Y_A = 21$   
 $T_A = 1$

B  $Y_B = 11$   
 $T_B = 0$

C  $Y_C = 9$   
 $T_C = 0$

D  $Y_D = 5$   
 $T_D = 1$



I don't know about the individual causal effects. BUT  
I can help you TEST your ideas / HYPOTHESES about them.

① Idea/H<sub>0</sub>:  $Y_i(1) = Y_i(0)$

② H<sub>0</sub> implies  $Y_i = Y_i(0)$

③ H<sub>0</sub> and Randomization

imply  $\leq$  equally likely  $t_0(T, Y)$  values.

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect ( $\tau$ )
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	? - 11 = $\tau_B$
C	0	9	?	9	? - 9 = $\tau_C$
D	1	5	5	5	5 - ? = $\tau_D$

④ Compare observed to implied by H<sub>0</sub> using a p-value

⑤ (Later) Show that this is a good test

