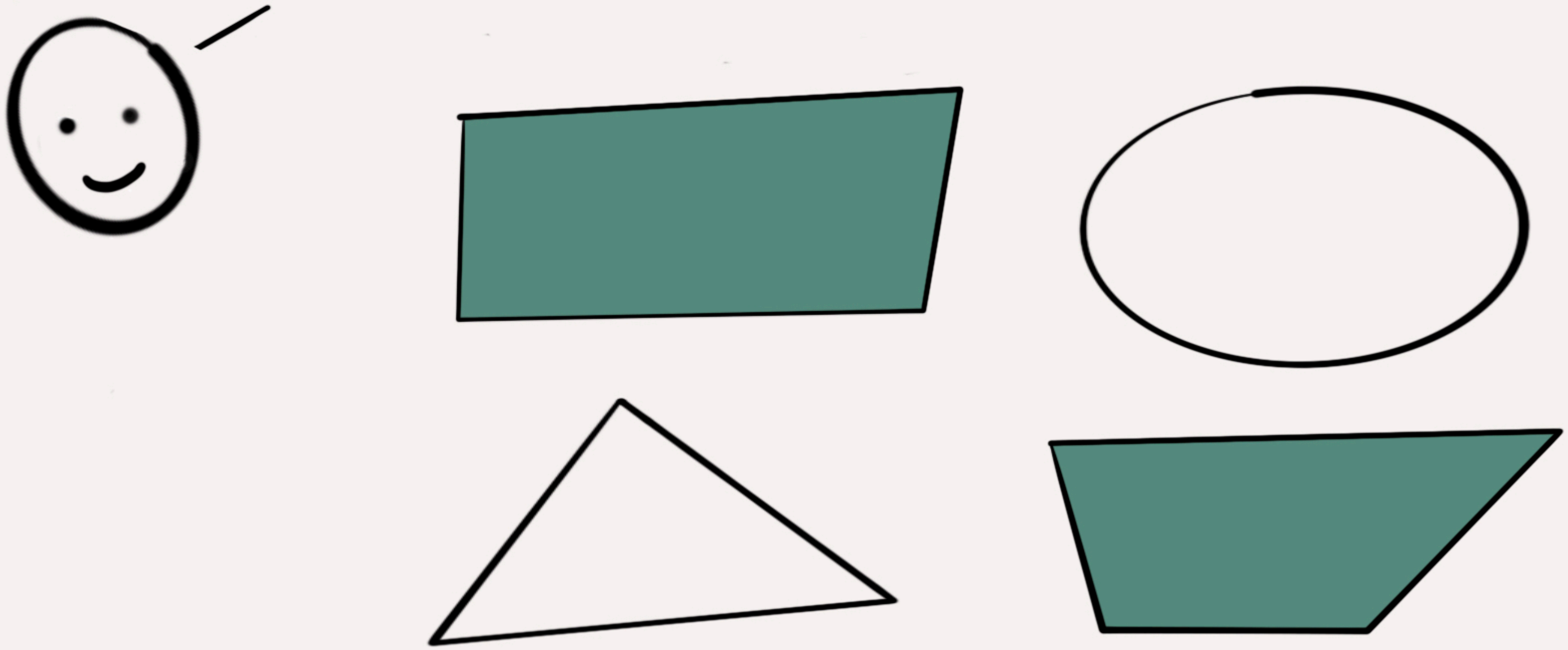
Statistics & Causal Inference:

Neyman & Fisher

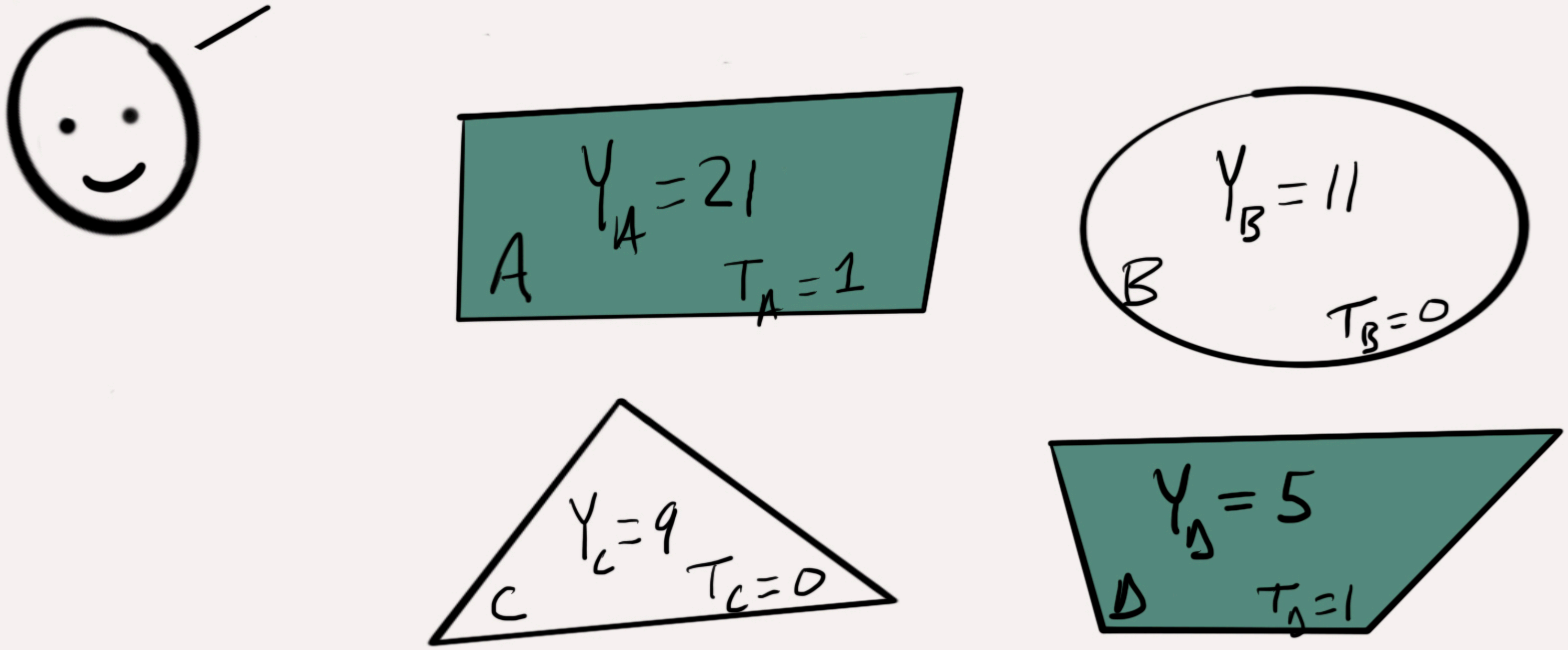
Jake Bowers
http://jakebowers.org

ICPSR 2021
19 July 2021

What is the causal effect?

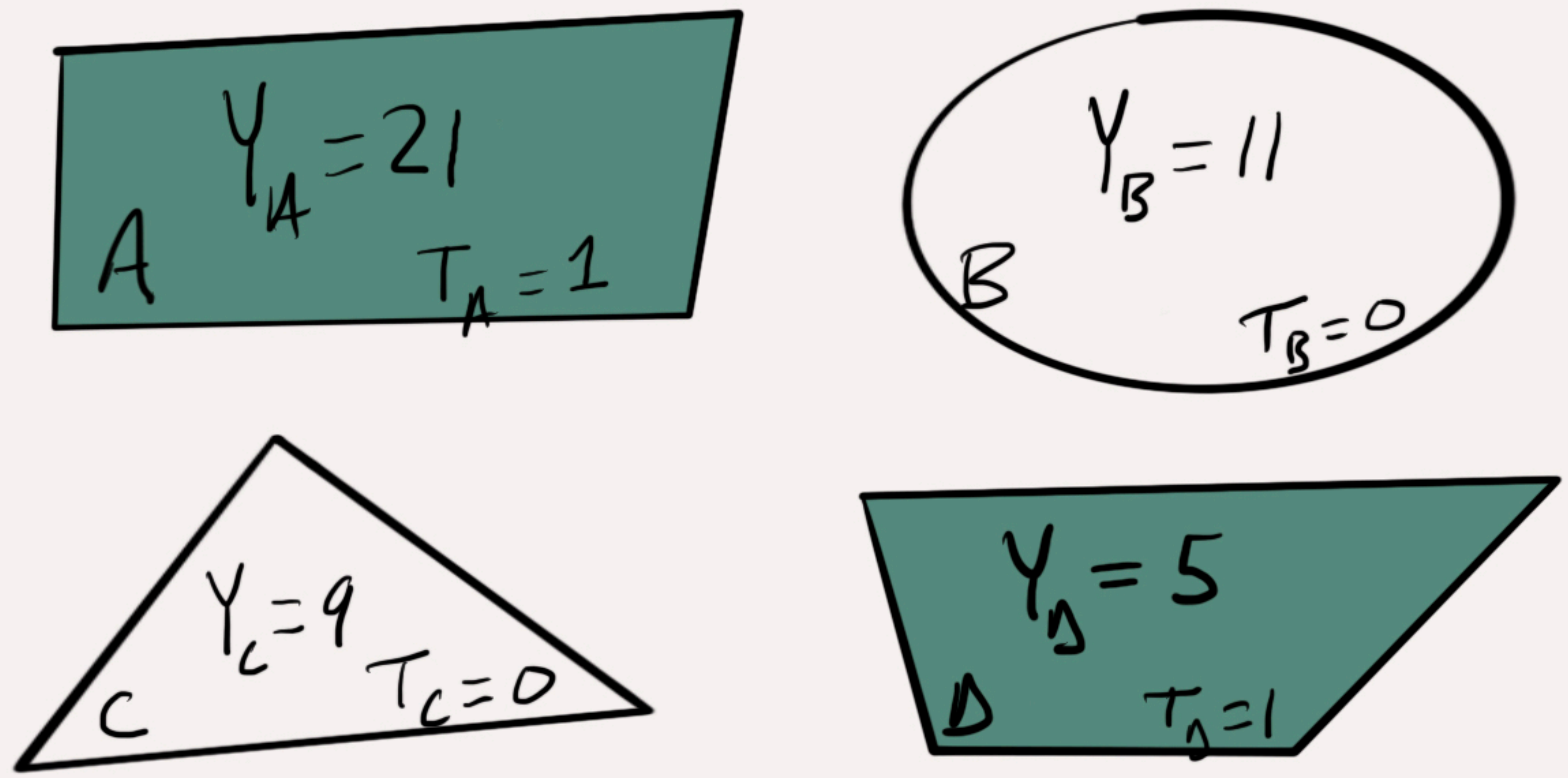


What is the causal effect?



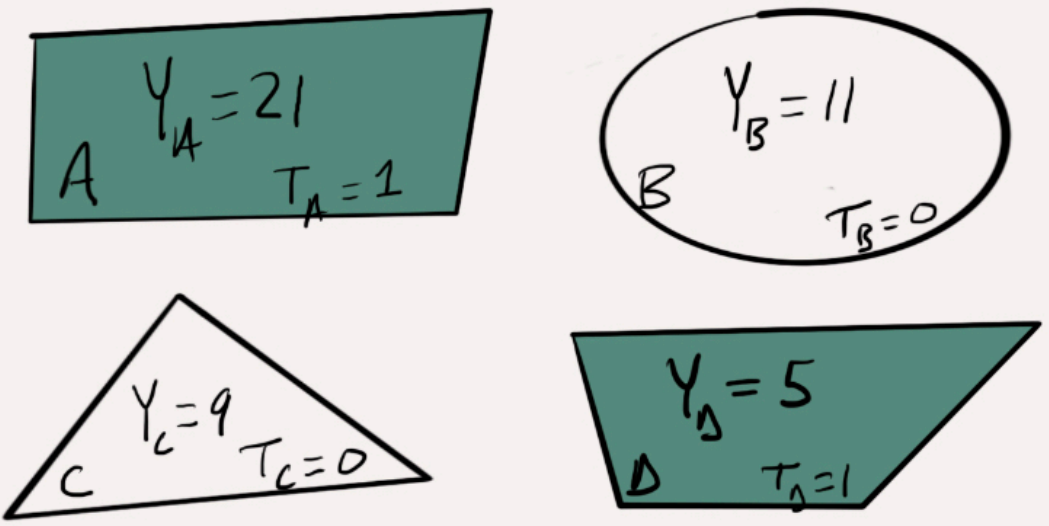
i Unit	T Treatment	Y Observed Outcome
A	1	21
B	0	11
C	0	9
D	1	5

What is the causal effect?



i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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	?	$5 - ? = \tau_D$

We don't know!



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

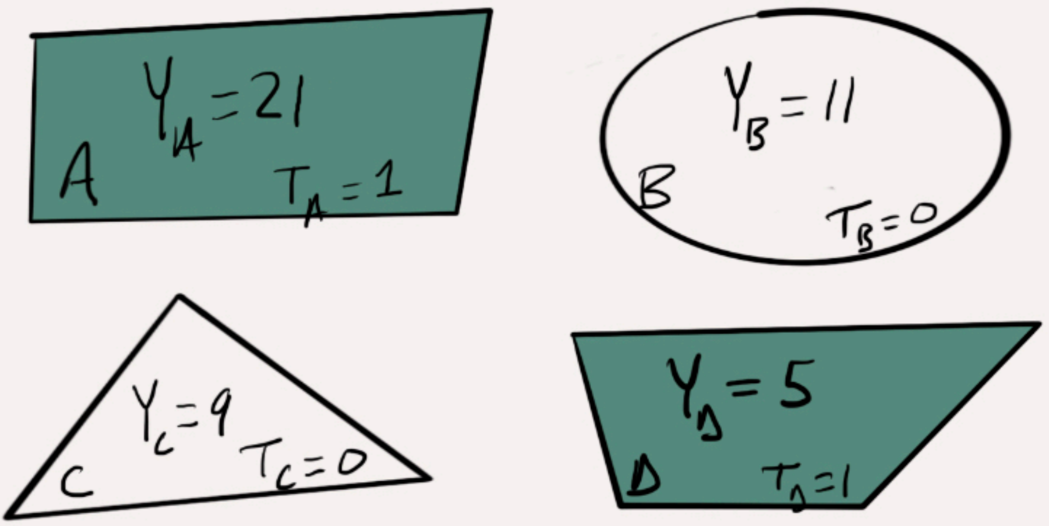
i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
A	1	21	21	?	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	?	$5 - ? = \tau_D$

Define: $\bar{\tau} = (\tau_A + \tau_B + \tau_C + \tau_D)/4$ (The avg. causal effect)

Notice: $\tau_A = Y_A(1) - Y_A(0)$, $\tau_B = Y_B(1) - Y_B(0)$, ...

$$\bar{\tau} = \left(\left(Y_A(1) - Y_A(0)\right) + \ldots + \left(Y_D(1) - Y_D(0)\right) \right)/4$$

$$= \left(Y_A(1) + \ldots + Y_D(1)\right)/4 - \left(Y_A(0) + \ldots + Y_D(0)\right)/4$$



I don't Know about the individual causal effects. <u>BUT</u> I can help you ESTIMATE the AVERAGE causal effect!

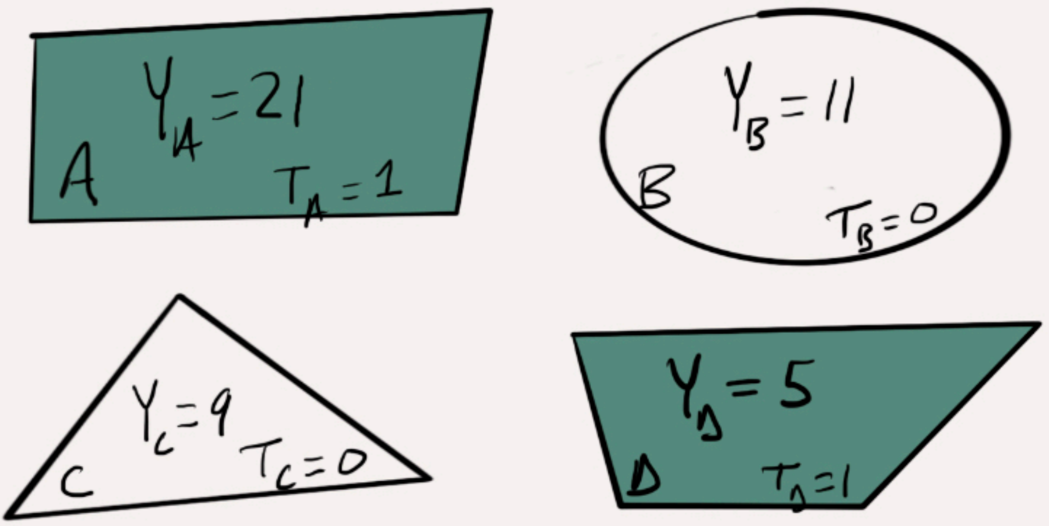
i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
A	1	21	21	?	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	?	$5 - ? = \tau_D$

Define: $\bar{\tau} = (\tau_A + \tau_B + \tau_C + \tau_D)/4$ (The avg. causal effect)

an Unobserved <u>Estimand</u>

$$= (Y_A(1) + \dots + Y_D(1))/4 - (Y_A(0) + \dots + Y_D(0))/4$$

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



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 (τ)
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	?	$5-?=\tau_D$

Define an Unobserved Estimand

$$\bar{\tau}=(\tau_A+\tau_B+\tau_C+\tau_D)/4$$

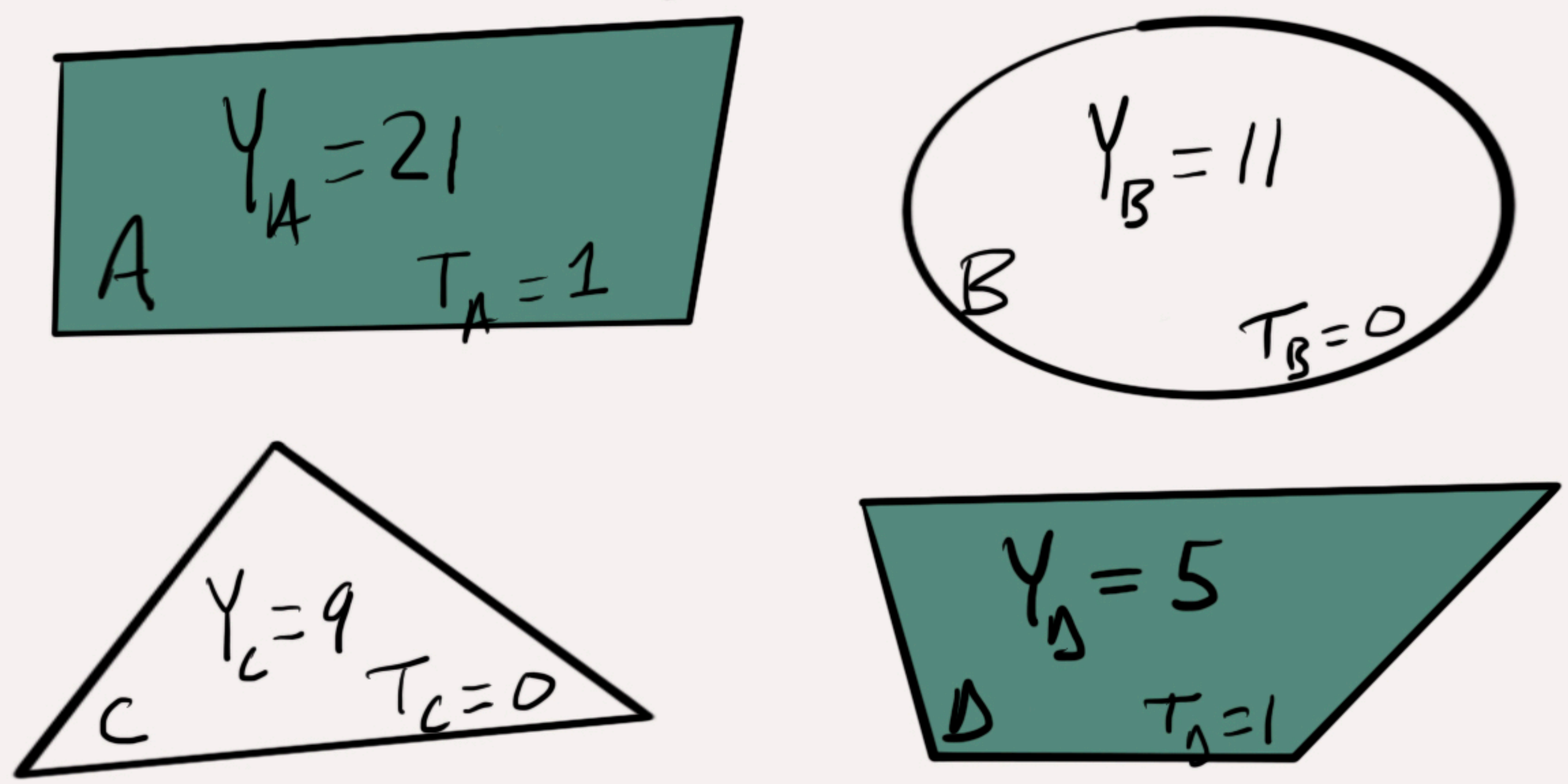
$$\bar{\tau}=\bar{Y}(1)-\bar{Y}(0)$$

Choose an Estimator

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

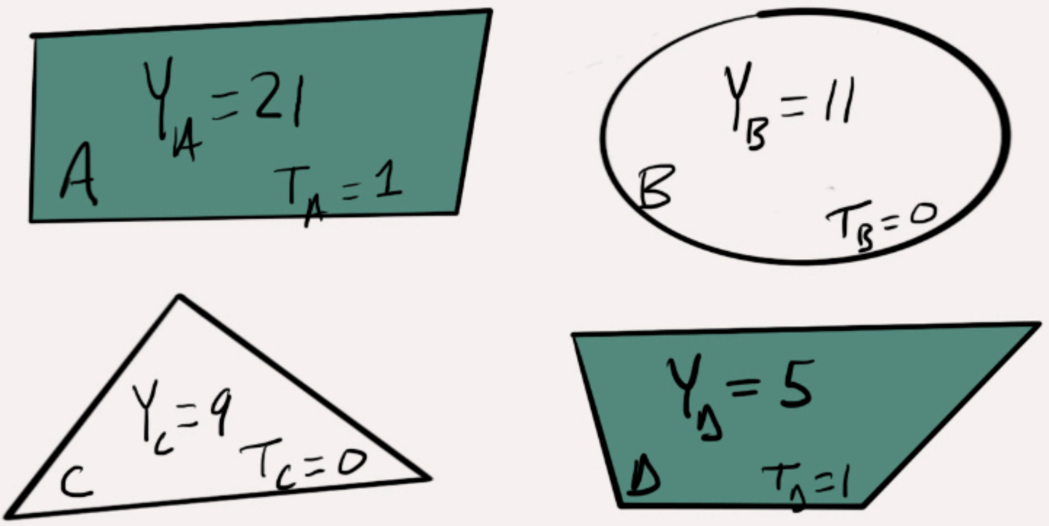
(Show (Later) that $\hat{\bar{\tau}}$ is a good estimator of $\bar{\tau}$)

What is the causal effect?



i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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	?	$5 - ? = \tau_D$

We don't know!

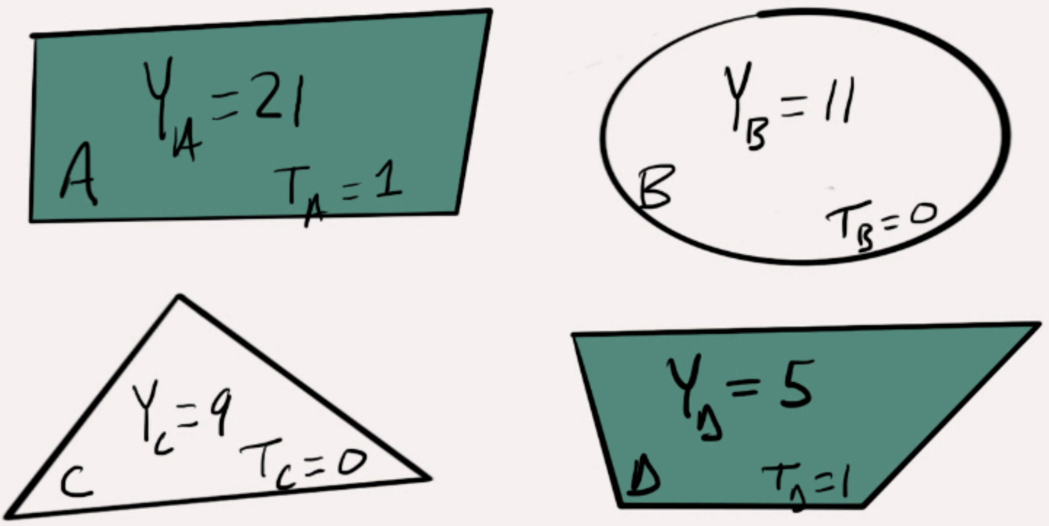


I don't know about the individual causal effects. BUT I can help you TEST your ideas/HYPOTHESES about them.

An idea: "No effects"

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

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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	?	$5 - ? = \tau_D$



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

An idea: "No effects"

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

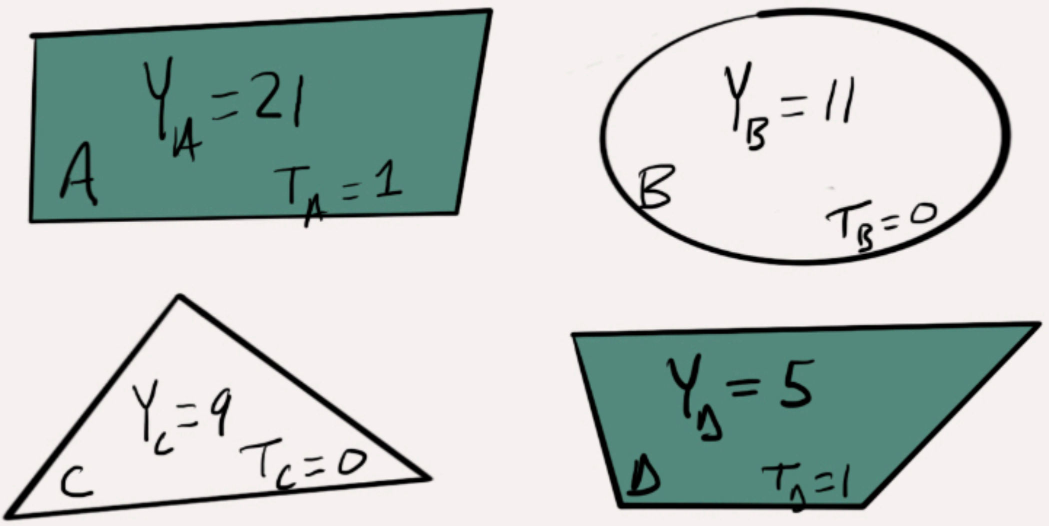
i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	5	$5 - ? = \tau_D$

What does this idea imply?

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

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

$= Y_i(0)$



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

An idea: "No effects"

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

$H_0 \Rightarrow \underset{\text{observed}}{Y_i} = Y_i(0)$

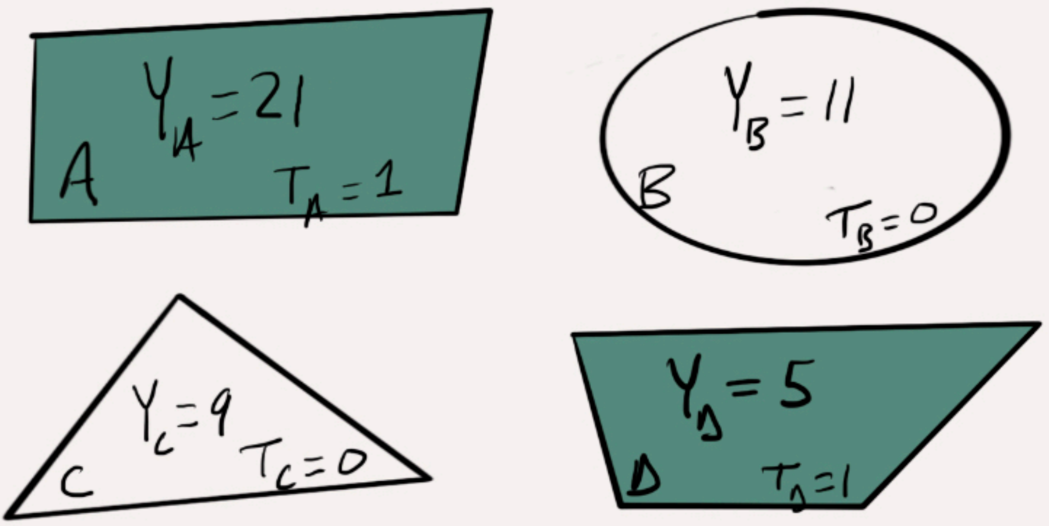
i	T	Y	Y(T=1)	Y(T=0)	Effect (τ)
A	1	21	21	21	$21 - ? = \tau_A$
B	0	11	?	11	$? - 11 = \tau_B$
C	0	9	?	9	$? - 9 = \tau_C$
D	1	5	5	5	$5 - ? = \tau_D$

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

But, A & D are only Treated by chance.

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

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



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 $Y_i = Y_i(0)$

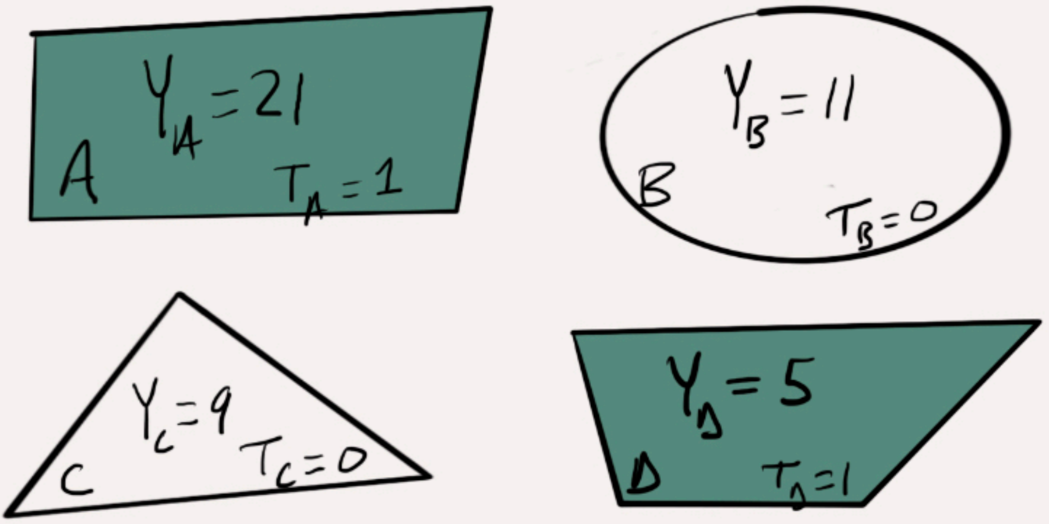
③ H_0 implies that a test statistic summarizing $T \to Y$,

$$t(T, Y) = \left(\frac{21+5}{2}\right) - \left(\frac{11+9}{2}\right) = 3$$

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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	$5 - ? = \tau_D$

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



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 $Y_i = Y_i(0)$

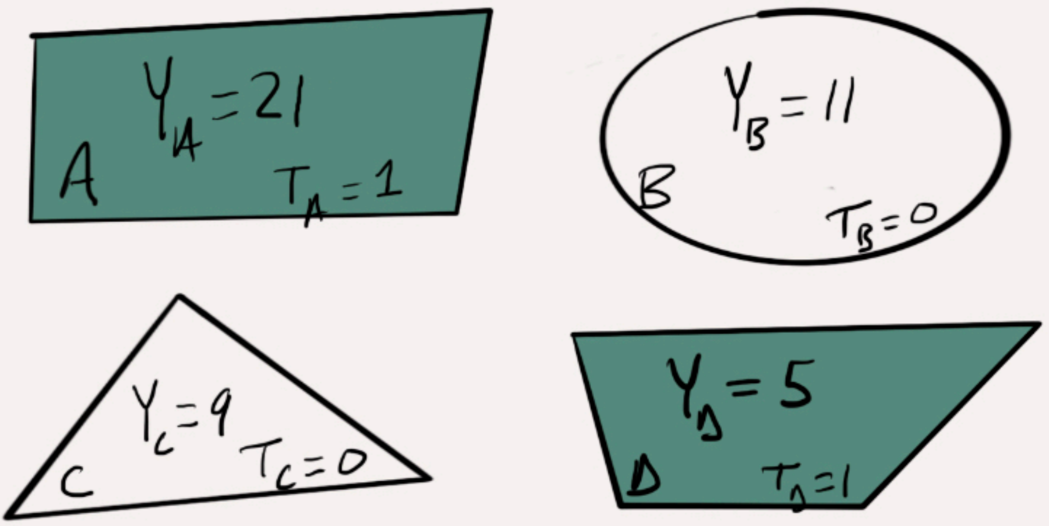
③ H_0 and Randomization imply 6 equally likely $t(T, Y)$ values.

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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
	3	9	7	−3	−7	−9

$t(T, Y)$ No effects



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

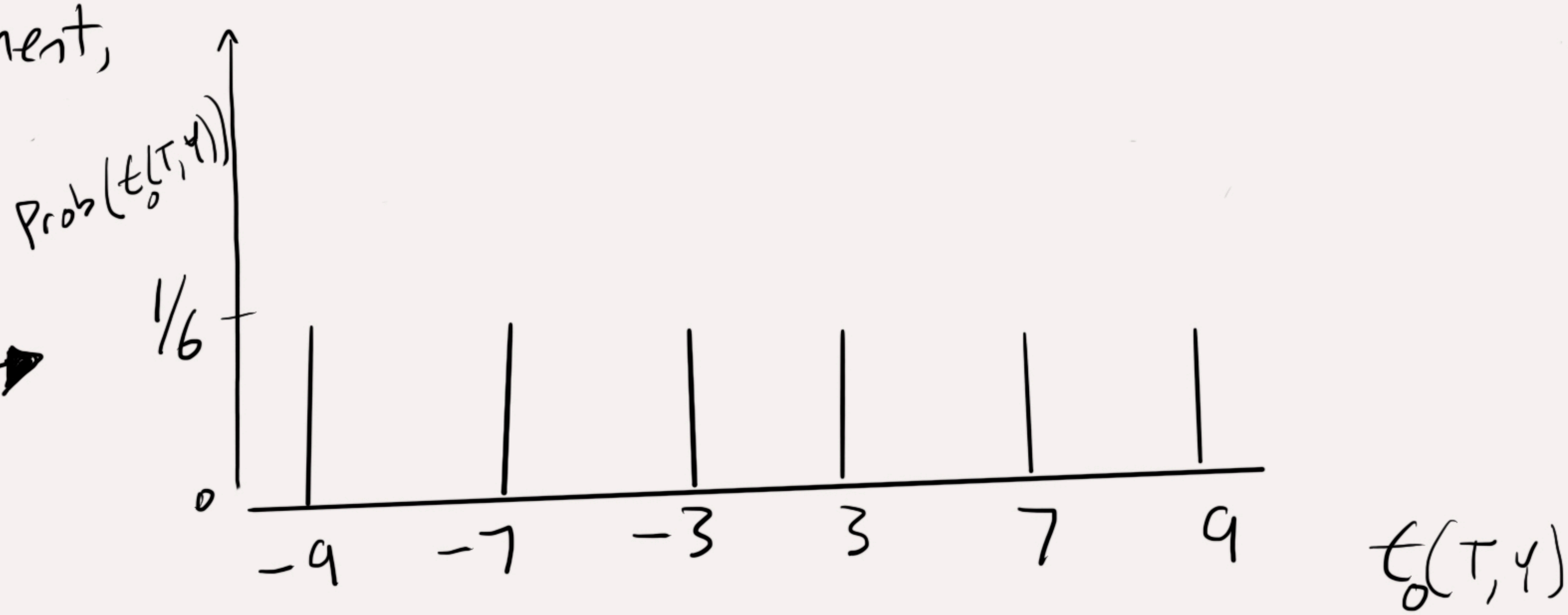
① Idea / H_0: $Y_i(1) = Y_i(0)$

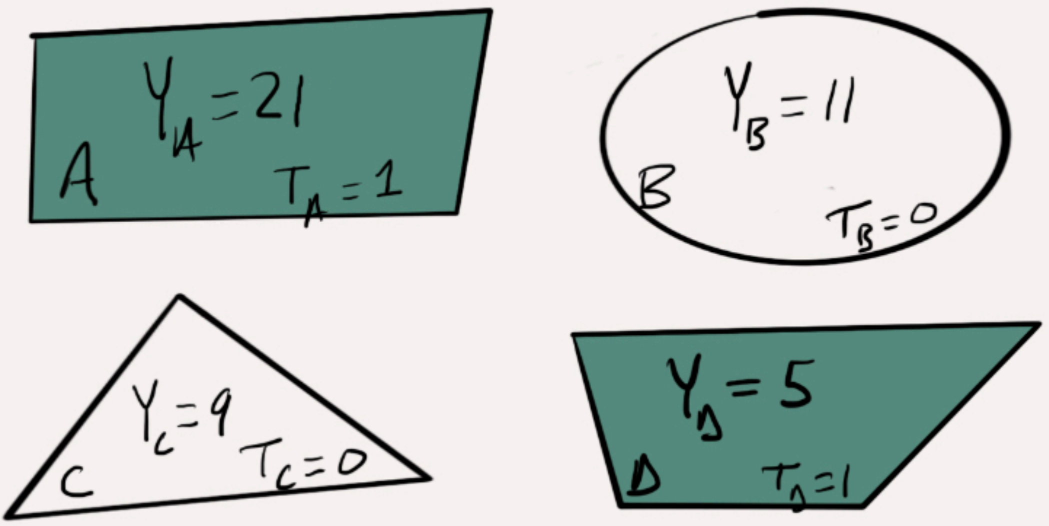
② H_0 implies $Y_i = Y_i(0)$

③ H_0 and Randomization imply <u>6</u> equally likely $t_0(T, Y)$ values.

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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



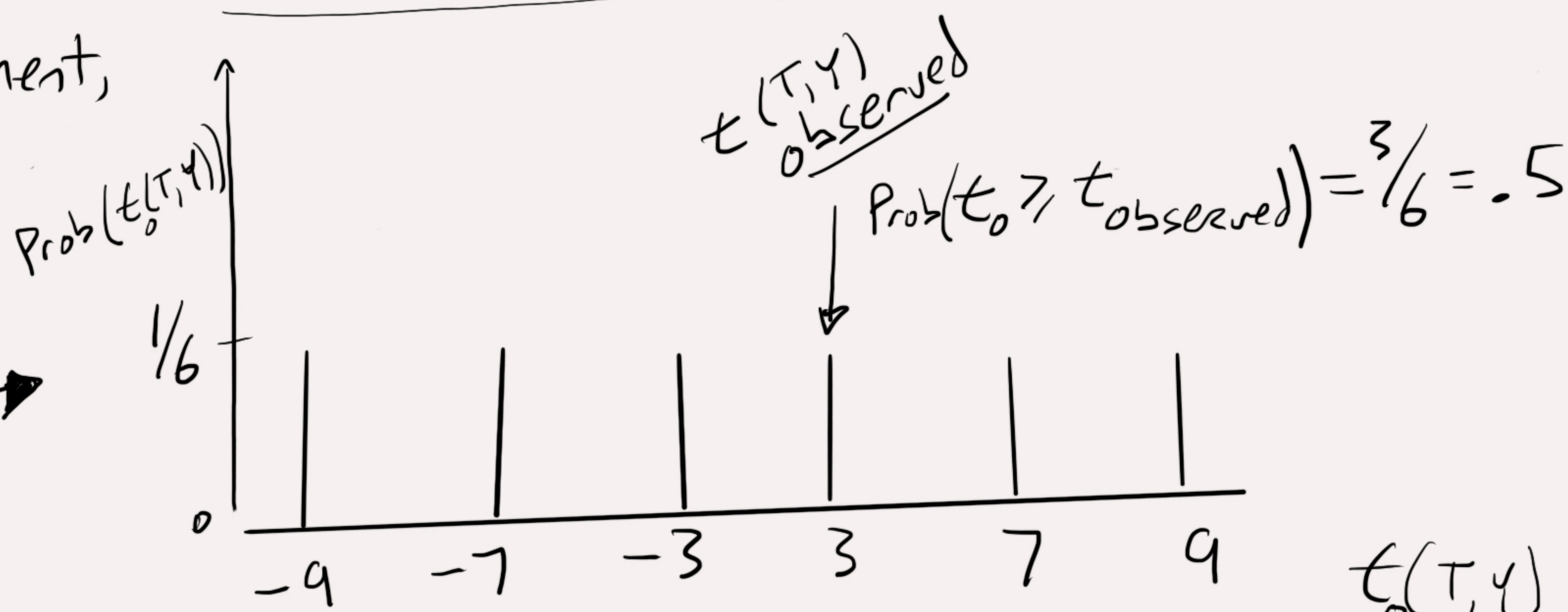


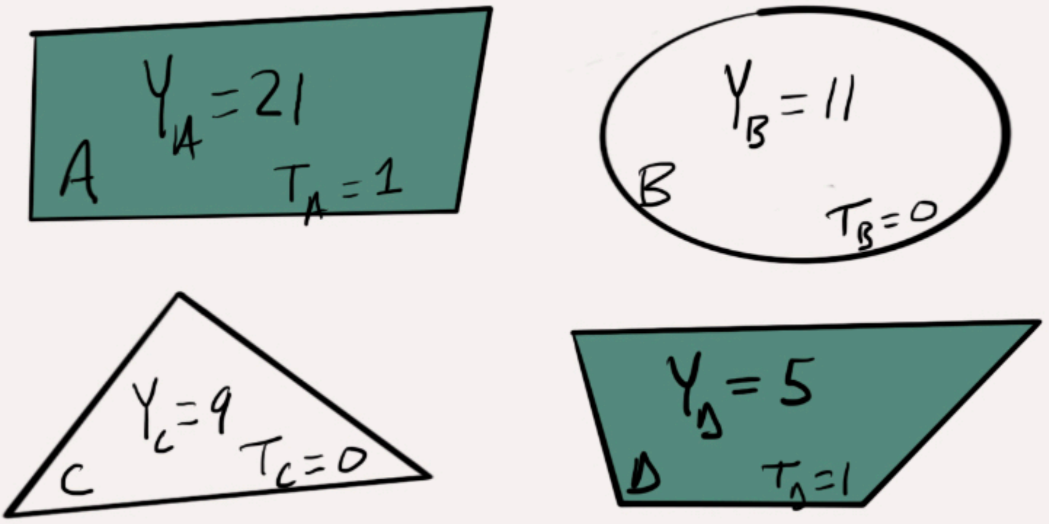
I don't know about the individual causal effects. BUT I can help you TEST your ideas/HYPOTHESES about them.

1. Idea/H_0: $Y_i(1)=Y_i(0)$
2. H_0 implies $Y_i=Y_i(0)$
3. H_0 and Randomization imply 6 equally likely $t_0(T,Y)$ values.

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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 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 and RAndomization imply 6 equally likely $t_0(T,Y)$ values.
4. Compare observed to implied by H_0 using a p-value
5. (Later) Show that this is a good test

i	T	Y	$Y(T=1)$	$Y(T=0)$	Effect (τ)
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	$5 - ? = \tau_D$