ONLINE MASTERS IN **DATA SCIENCE**

DSC 215 - PROBABILITY AND STATISTICS FOR DATA SCIENCE

DIFFERENCE OF MEANS

POWER CALCULATIONS

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

Definition of Statistical Power

- **Definition:** The statistical power of a hypothesis test is the probability that the test correctly rejects the null hypothesis when a ***specific*** alternative hypothesis is true.
- Roughly speaking, it is represents the chances of a "true positive" detection conditional on the actual existence of an effect.
- Mathematically:

$$\boxed{\text{power} = \mathbb{P}\left(\text{reject } H_0 \mid H_1 \text{ is true}\right)}$$

- From the definition: $\text{power} = 1 - \mathbb{P}(\text{Type II Error})$

Why Statistical Power is Important

- **Recall:**

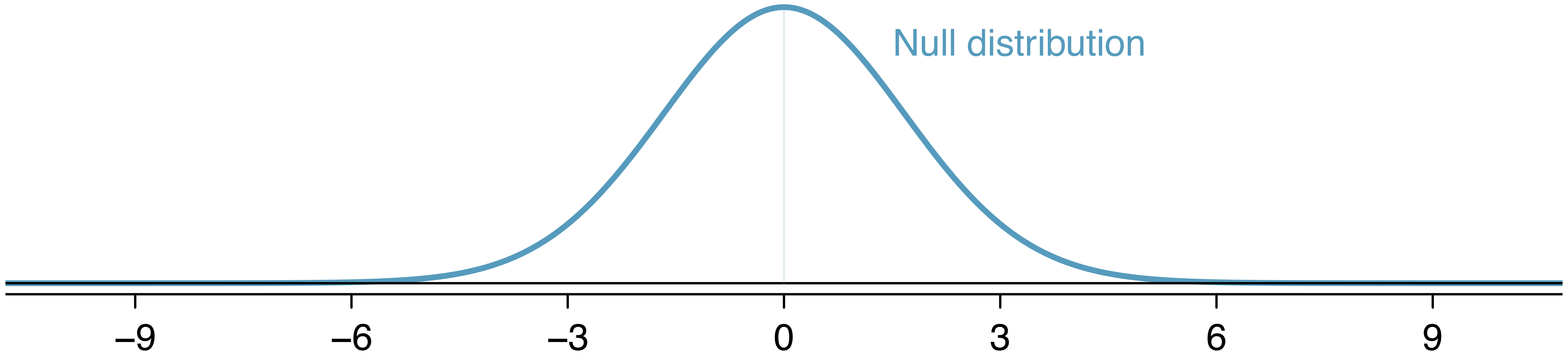
$$\boxed{\text{power} = \mathbb{P}\left(\text{reject } H_0 \mid H_1 \text{ is true }\right)}$$

- Consider the example of a blood-pressure drug trial, where
  - H_0: the drug has no effect, i.e., $\mu_t - \mu_c = 0$
  - $H_1 : \mu_t - \mu_c = -3mmHg$
- When designing the trial, you would like to know that
  - **If** the drug actually has an effect of some size (e.g., reduces blood pressure by 3mmHg),
  - **then** the probability of rejecting the null is high.
- How can we make sure of this? We make sure the power is high enough.

Statistical Power: an Example

- Suppose for the study we would like to run, $n_t = n_c = 100$, and we have *a previous estimate from other studies of* $s_t = s_c = 12$.
- Then we can estimate $SE_{\bar{x}_t - \bar{x}_c} = \sqrt{12^2/100 + 12^2/100} = 1.7$
- Moreover, since $df > 30$

  under the null, we can model $\bar{x}_t - \bar{x}_c$ as normal, with mean 0 and SE = 1.7



Calculating the Statistical Power

- How can we make sure the power is high enough?
- $\mathbb{P}\left(\text{reject } H_0 \mid H_1 \text{ is true}\right) = \mathbb{P}\left(p\text{-value} < \alpha \mid H_1 \text{ is true}\right)$
- p-value $< \alpha \iff \mathbb{P}(\text{test statistic at least as extreme as observed} \mid H_0) < \alpha$

  $\iff \bar{x}_t - \bar{x}_c$ lies in the rejection region

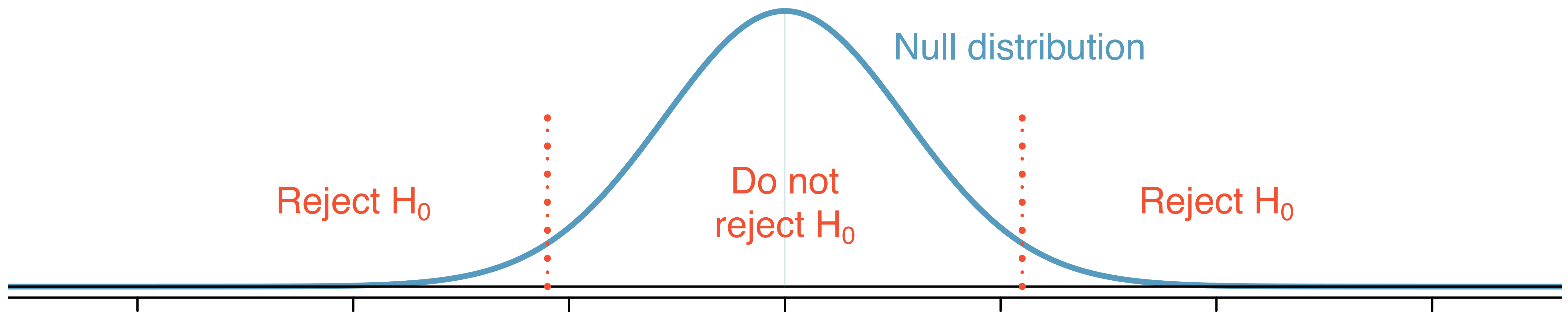
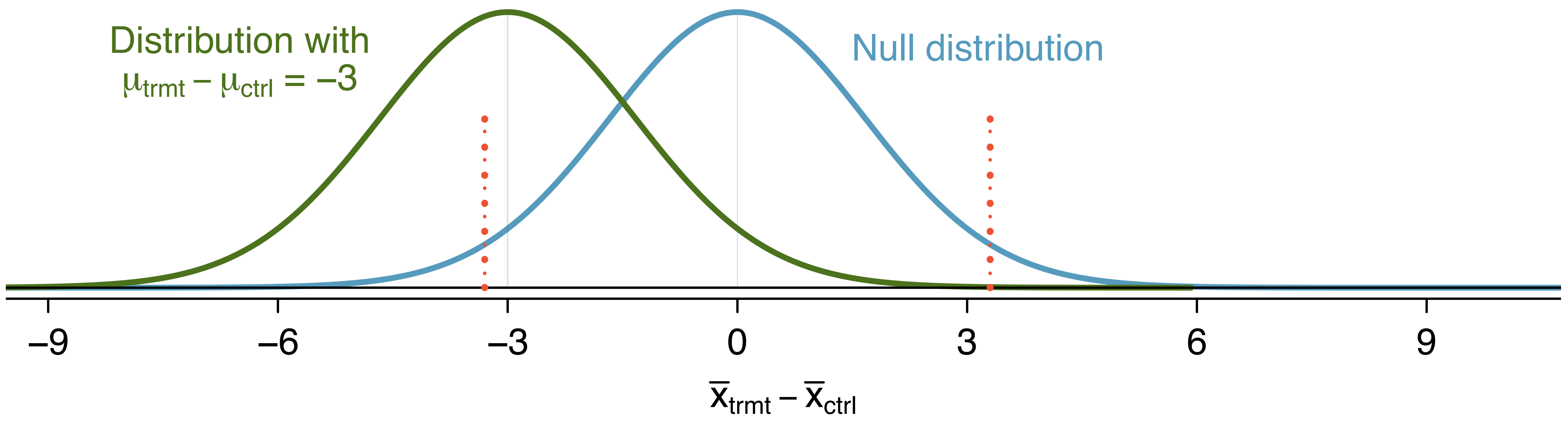


Figure: Distribution of $\bar{x}_t - \bar{x}_c$

- For example, $\alpha = 0.05$, means we would reject H_0 if $\bar{x}_t - \bar{x}_c$ is less than $-1.96 \times SE = -3.332$ or greater than $1.96 \times SE = 3.332$

Calculating the Statistical Power

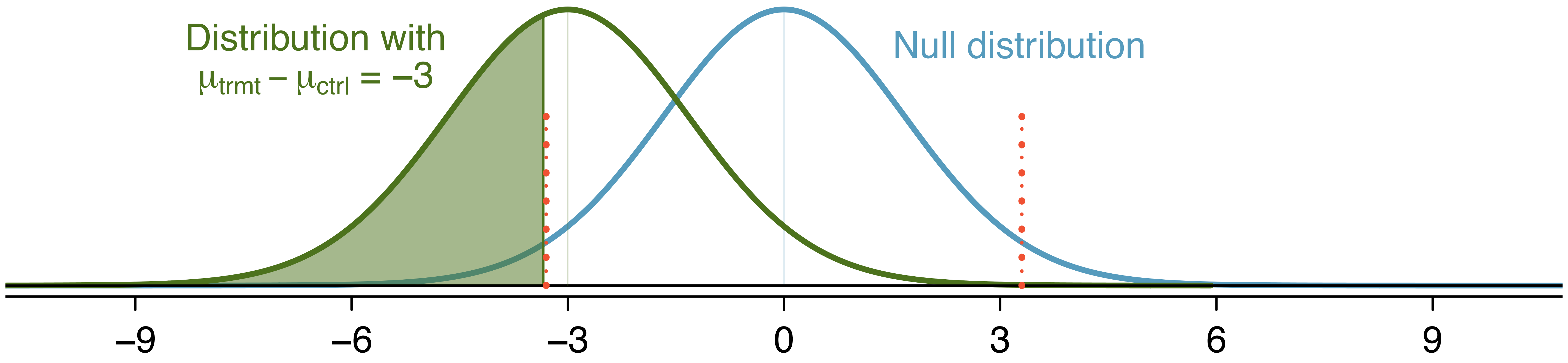
- But how do we use this to calculate the power?
- **What we calculated:** the boundaries of the rejection region under the null
  - In our example: ± 3.332
- **What we know**: The alternate distribution is normal with mean -3mmHg (recall $H_1 : \mu_t - \mu_c = -3mmHg$)



- **What remains:** calculate the probability, under H_1, of falling in the rejection region.

Calculating the Statistical Power

- **What remains:** calculate the probability, under H_1, of falling in the rejection region.



- This is the probability under a normal distribution of mean -3, and SE 1.7, of drawing $x < -3.332$.
(Here $df > 30$, so we can approx. with a normal distribution)

- Using software or Z-scores: $Z = \dfrac{-3.332 - 3}{1.7} = -0.2 \implies P(x < -3.332) \approx 0.42$

Ensuring the Statistical Power is High

- In the last example, we found power $= \mathbb{P}\left(\text{reject } H_0 \mid H_1 \text{ is true}\right) = 0.42.$
- So in this example, there is **only** a 42% chance of rejecting the null hypothesis when the alternate hypothesis holds!
- This is clearly not good.
- **Question:** What could we have done to ensure higher power?
- **Answer:** Pick larger n!
- *Repeating the same example with $n_t = n_c = 500$, gives rejection boundaries ± 1.49, $SE = 0.76$, Z-score under the alternate $Z = 1.99$, and power = 0.977.*

Working Our Way Backwards to Find a Good n

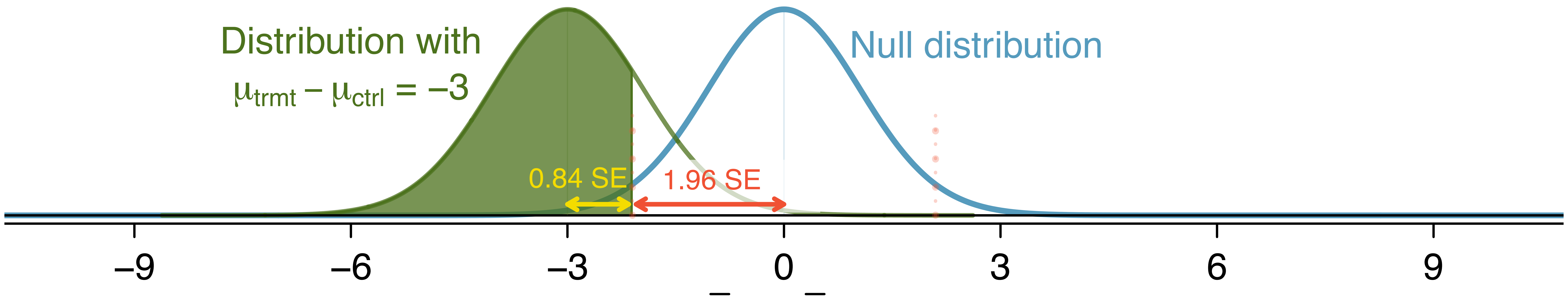
- We've calculated $n_t = n_c = 100 \implies$ power $= 0.42$

  $n_t = n_c = 500 \implies$ power $= 0.977$

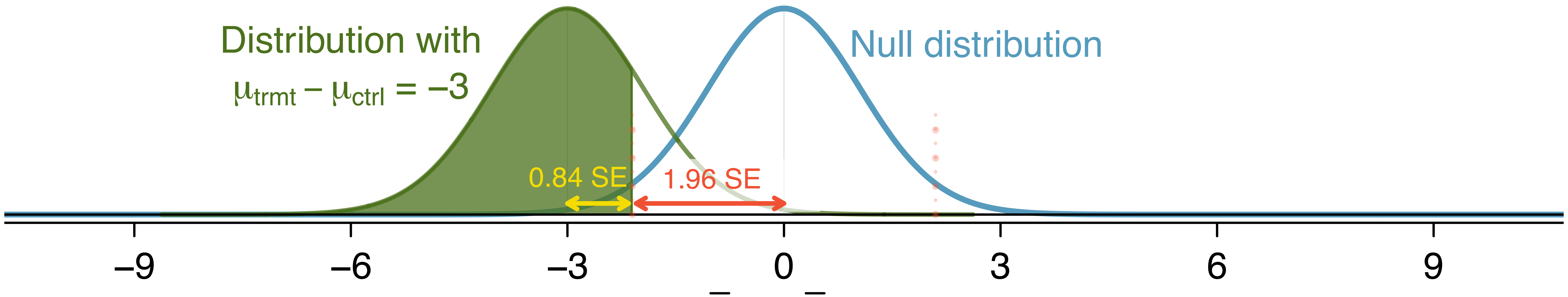
- **Question:** If we want power $= 0.8$, what should n be?

- **Approach:**

  - First, find Z-score associated with a lower tail of 80% $\implies Z = 0.84$
  - Second, at $\alpha = 0.05$, rejection boundary is $-1.96 \times SE$
  - So the distance between the means under the null and the alternate is $1.96 \times SE + 0.84 \times SE = 2.8 \times SE$



Working Our Way Backwards to Find a Good n



- So, now we know

$$2.8 \times SE = 0 - (-3) = 3$$

$$2.8 \times \sqrt{12^2/n + 12^2/n} = 3$$

$$\implies n = 250.88$$

- We should aim for 251 patients per group to achieve 80% power at $\alpha = 0.05$