Learning Objectives

- **Describe** the four possible decisions in the *Null Hypothesis Statistical Testing (NHST)* framework
- **Identify** factors that increase the likelihood of rejecting the null hypothesis (H_0)
- **Justify** decisions to manage error
- **Visualize** statistical power

Announcements

- Midterm exam reviews ongoing
  - See corrected item in announcements
  - Even if you missed the midterm
- SPSS/Jamovi 4 is available (Sign Test)
- No video for Monday
- Ch. 10 is <u>important</u> (for Ch. 11 focus on lecture)

Inferential Statistics

- Quantify <u>confidence</u> in whether our pattern of observations will replicate
  - **Goal**: Rule <u>out</u> random chance as potential cause
    - a.k.a, reject H_0
    - “My estimates will fluctuate but <u>the pattern is reliable</u>; (it would replicate if I sampled again)”

Possible conclusions:

1. We can rule out random chance
    - “The relationship is ***probably*** real”

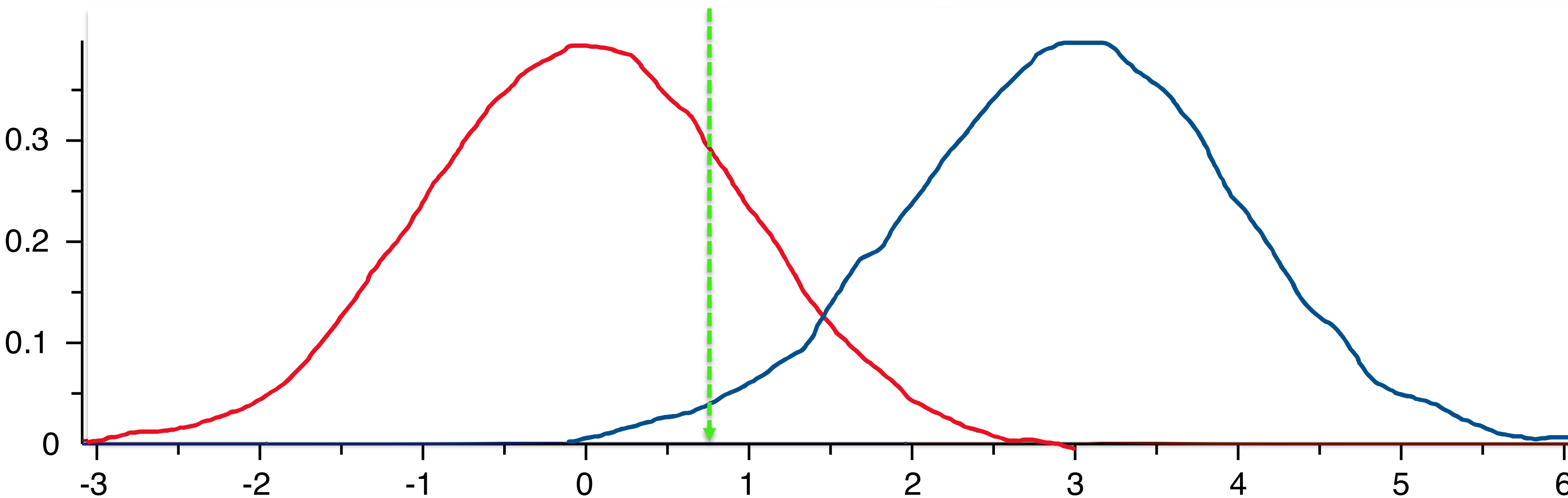
2. We cannot rule out random chance*
    - “The relationship ***could easily*** be random”

Note:** We do not conclude that chance <u>**is the cause**</u>, but it ***could be

Inference: Which distribution did our observation come from?

Distribution$|H_0$: Red curve
Distribution$|H_1$: Blue curve

Decision threshold
(we choose this value)



What happiness scores look like if treatment did not work

What happiness scores look like if treatment worked

Truth/Reality

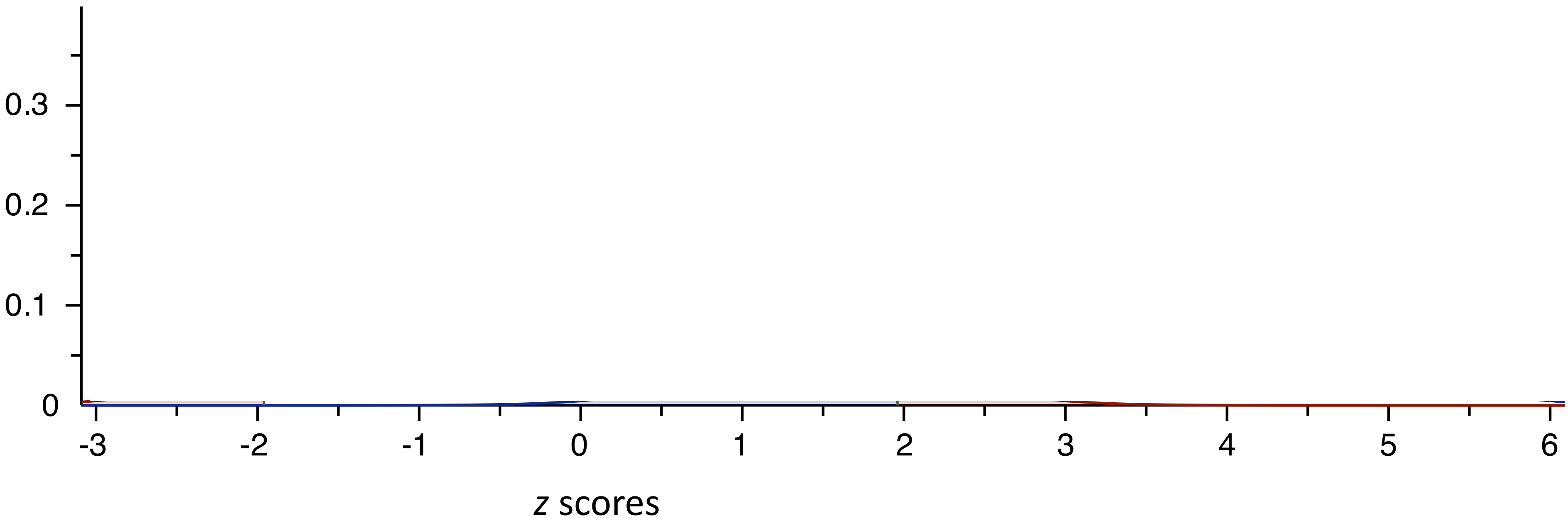
Our Decision	**No effect:** Relationship is random	**Real effect:** Relationship is real
Don’t reject chance: “Relationship could easily be random”	**Correct!** $(1 - \alpha)$	**Type II Error** β
Reject chance: “Relationship seems real”	**Type I Error** α	**Correct!** $(1 - \beta)$

Inference: Which distribution did our observation come from?

	No effect	Yes effect
Don't reject H_0	**Correct!** $(1 - \alpha)$	**Type II Error** β
Reject H_0	**Type I Error** α	**Correct!** $(1 - \beta)$

Distribution|H_0: Red curve
Distribution|H_1: Blue curve

Chosen value for α



Alpha and Beta

Beta (β) is the probability of **Type II error**

- We say "there is no effect", but there <u>is</u> an effect

Alpha (α) is the probability of **Type I error**

- We say "there is an effect", but there <u>is not</u> an effect

Mathematical Formulae

When H_0 is true, <u>in reality there is **no effect**</u>

- We reject H_0, a *Type I* error, <u>or</u>
- We correctly retain H_0

$$\boxed{1 = \alpha + (1 - \alpha)}$$

When H_0 is not true, <u>in reality there **is an effect**</u>

- We retain H_0, a *Type II* error, <u>or</u>
- We correctly reject H_0

$$\boxed{1 = \beta + (1 - \beta)}$$

Error Management

- **Goal:** Be correct all the time!
  - Impossible goal

- **Reality:** <u>Manage errors</u> in a systematic way

- Which errors are worse? Type I or Type II?
  - **Verdict:** Guilty vs. not guilty
  - **Medicine:** New Coronavirus test
    - Type I error: Test says you are sick, but you aren’t sick
    - Type II error: Test says you aren’t sick, but you are

Controlling Error rates

Controlling Type II (or β) errors:

- Type II errors less likely with more <u>statistical power</u> (1- β)

more likely to reject the null hypothesis

  - **Larger *N***
  - **Smaller $\sigma_{\mathbf{pop}}$**
  - **Larger difference between means**
  - **Larger α, or 1-tail test**

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

$$z_{\mathrm{obt}} = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

When do we reject the null?

- Reject when z_{obt} gets large enough
  - Specifically when ± 1.96 (for 2-tail tests)
    - Because z_{crit} is always ± 1.96 or ± 1.645 (for 1-tail)
    - Critical values change for other test statistics (e.g., *t*, *F*)

- **Larger *N*** increases z_{obt} (because it reduces $\sigma_{\bar{X}}$)
- **Smaller $\boldsymbol{\sigma_{\text{pop}}}$** increases z_{obt} (b/c it reduces $\sigma_{\bar{X}}$)
- **Larger difference** $(\bar{X} - \mu)$
- **Larger $\boldsymbol{\alpha}$**, b/c it reduces z_{crit}

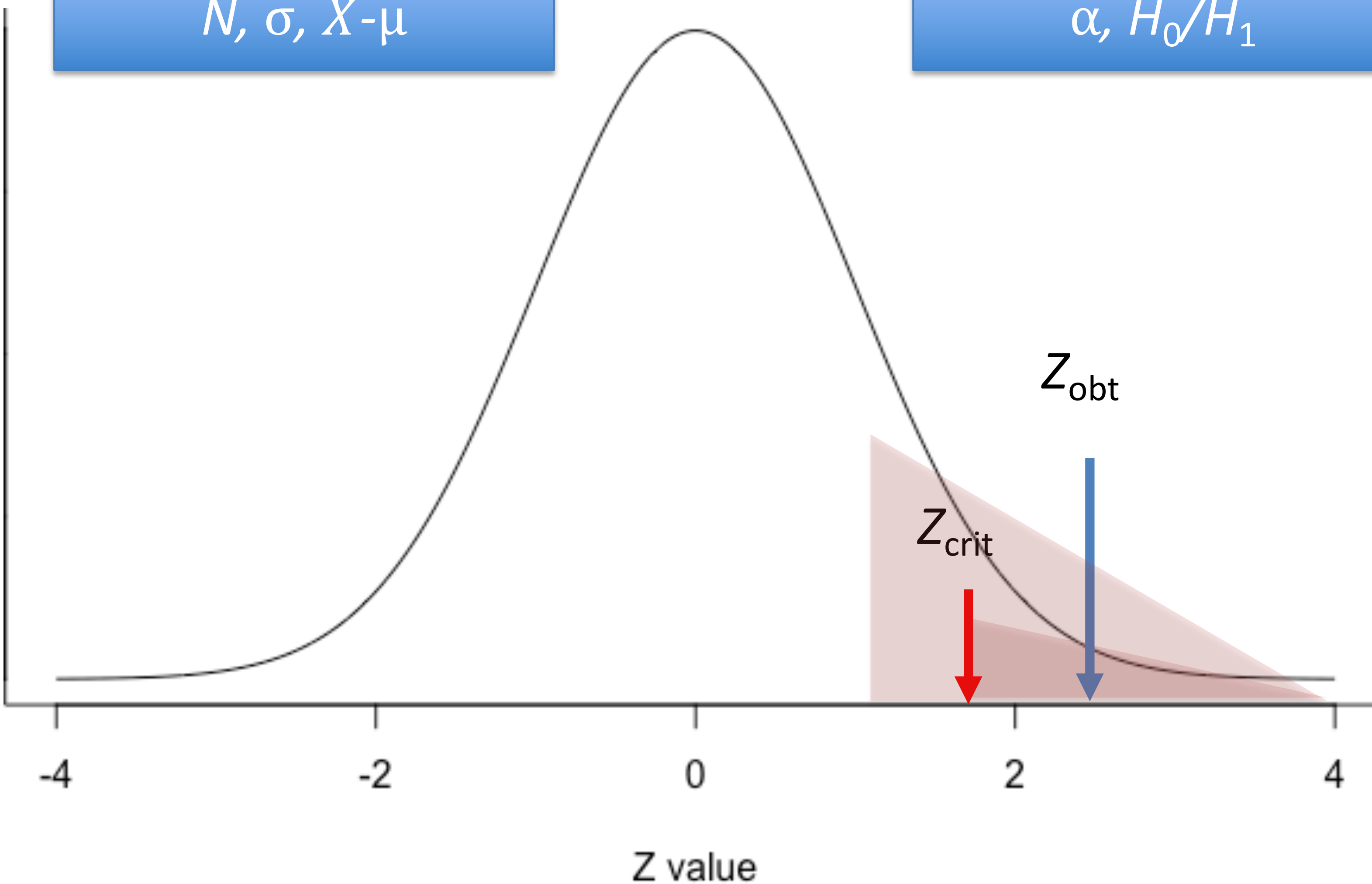
  (z_{obt} does not need to be as extreme to reject H_0)
- **Directional *H_0/H_1***, b/c it reduces z_{crit}

  (z_{obt} does not need to be as extreme to reject H_0)

z_{obt} & z_{crit} are the moving parts:

Z_{obt} influenced by:
N, σ, $\bar{X}$-μ

Z_{crit} influenced by:
α, H_0/H_1



Type I vs. Type II errors

Reducing Type I errors <u>increases</u> Type II errors

Note: Reducing Type II errors does <u>not</u> increase Type I errors

- We need to balance errors:
  - What are the consequences of **failing to find a real effect?**
  - What are the consequences of **“finding” an effect that is not real**?

<u>We choose</u> acceptable rates of Type I error:

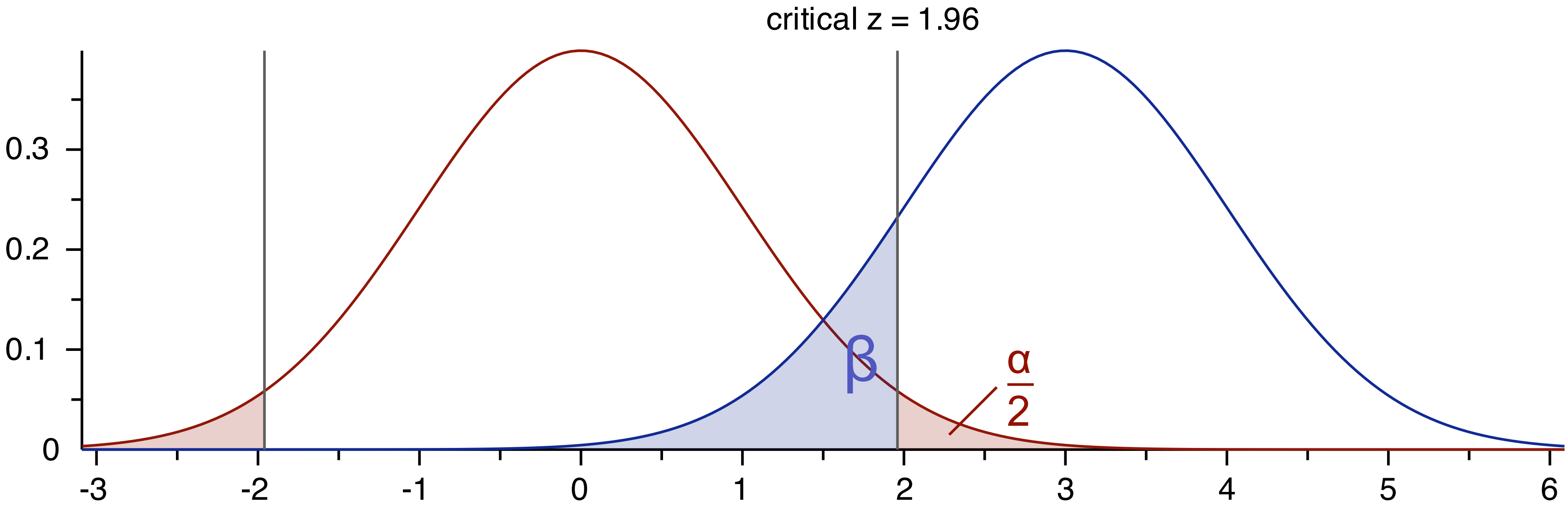
- Psychology: $\alpha = .05$
- Physics: $\alpha = .00000003$
- Biology (macro): $\alpha = .05$
- Biology (micro): $\alpha = .000001$

$(1-\beta)$ = **Power**

- Research seeks at least 80% power
  - Not 40-60%!!

- Visualize in ***G*Power*** [*(or Laken's Shiny)*]
  - Most common tool for calculating power
  - Normal distribution (or *z*-distribution)
    - What is/are the *z*-score for: $\alpha_{2\text{-tail}} = .05$

Red = H_0 Distribution (If H_0 were true)
Blue = H_1 Distribution (If H_1 were true)



$\alpha_{2\text{-tail}} = .05$

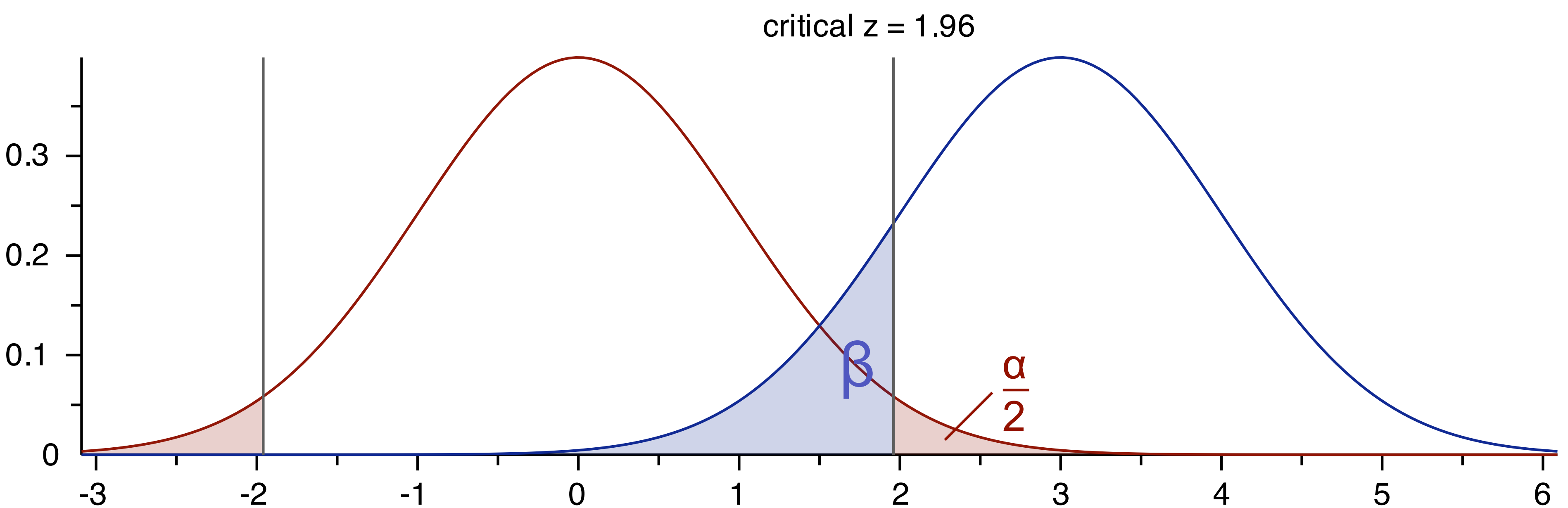
$\beta = .15$

Power = .85

$\bar{X}$-μ = 3 units

$\sigma_{\bar{X}}$ = 1 unit

Red = If H_0 is true
Blue = If H_1 is true



$\alpha_{2\text{-tail}} = .05$

β = .15

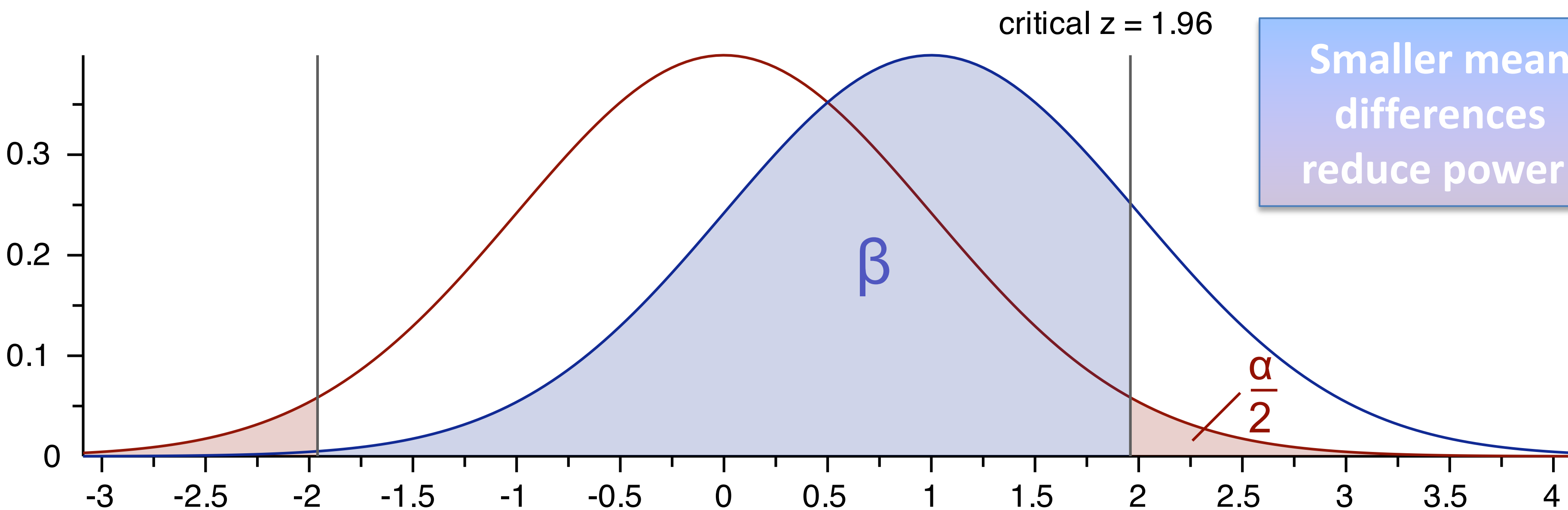
Power = .85

$\bar{X}$ - μ = 3 units

$\sigma_{\bar{X}}$ = 1 unit

Power when effect size is smaller!

Red = If H_0 is true
Blue = If H_1 is true



Smaller mean differences reduce power!

$\alpha_{2\text{-tail}}$ = .05

β = .83

Power = .17

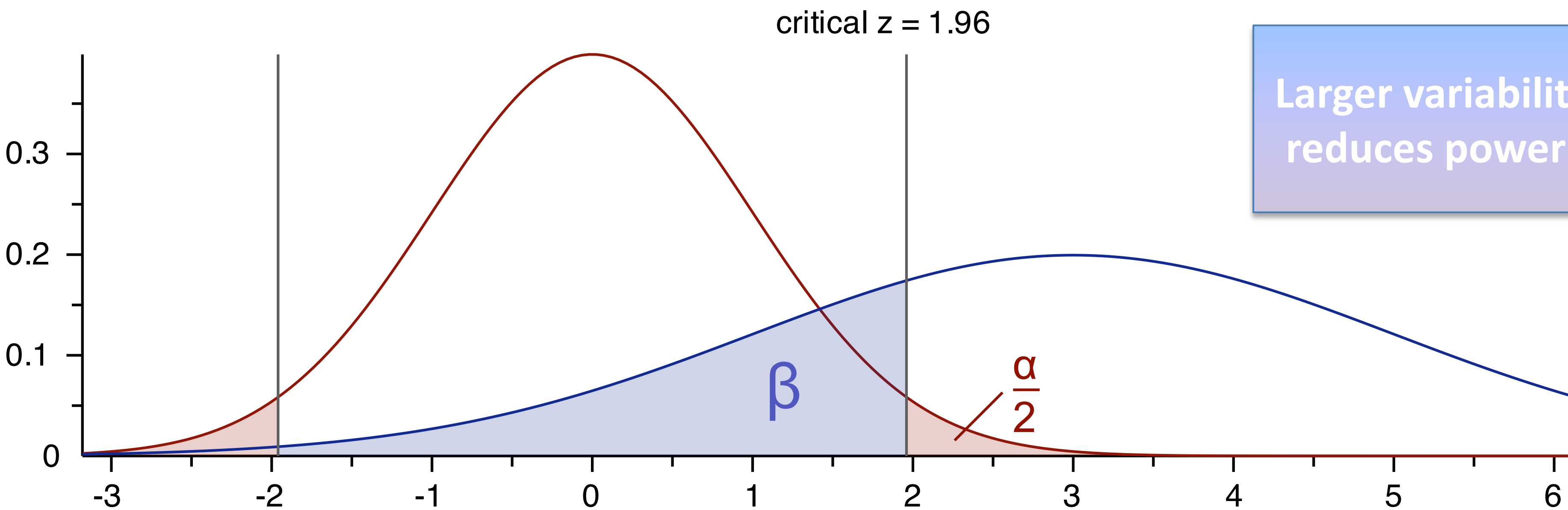
$\bar{X}$ - μ = **1 unit** **(was 3)**

$\sigma_{\bar{X}}$ = 1 unit

Power when variability is greater!

Red = If H_0 is true

Blue = If H_1 is true



Larger variability reduces power!

$\alpha_{2\text{-tail}} = .05$

$\boldsymbol{\beta}$ **= .29**

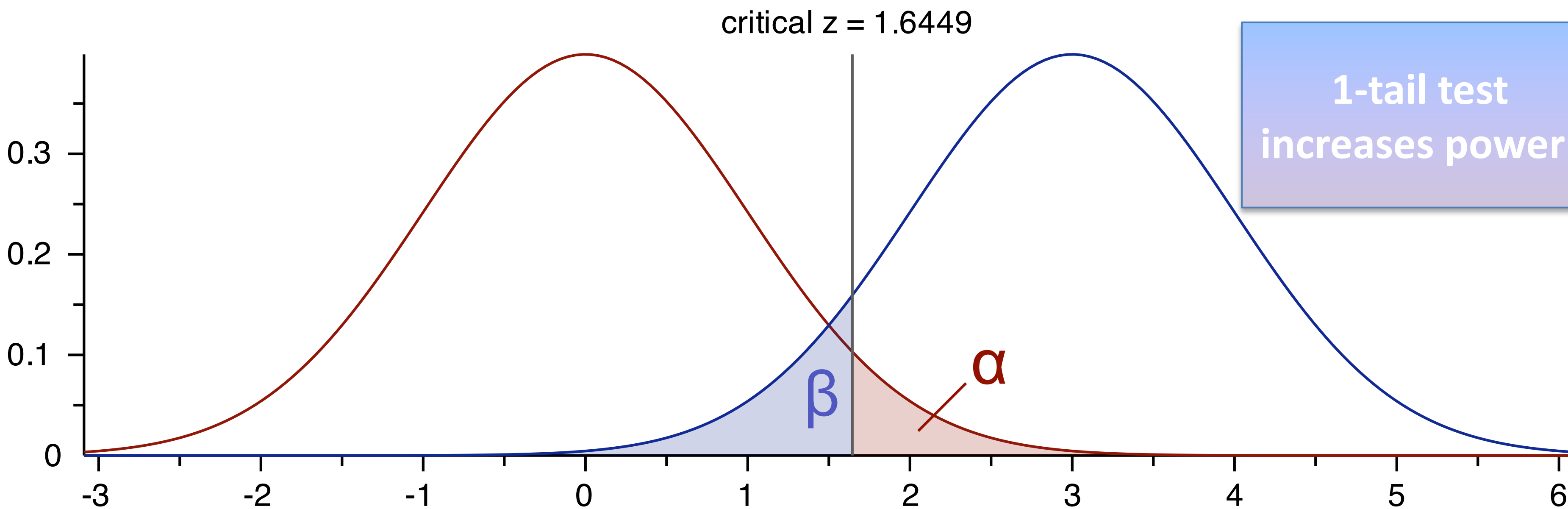
Power = .71

$\bar{X}$ - μ = 3 units

$\sigma_{\bar{X}}$ = **2 units** **(was 1)**

Power when 1-tail (vs. 2-tail) test!

Red = If H_0 is true
Blue = If H_1 is true



1-tail test increases power!

$\alpha_{1\text{-tail}} = .05$

$\beta = .09$

Power = .91

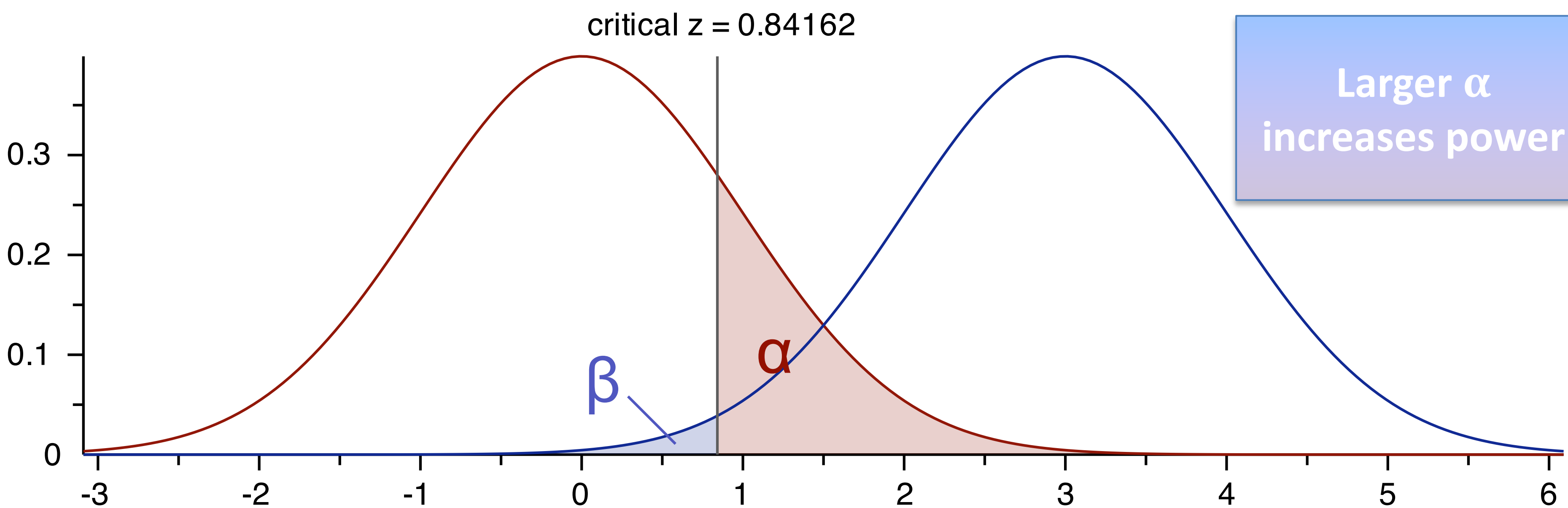
$\bar{X} - \mu = 3$ units

$\sigma_{\bar{X}} = 1$ unit

Power when α is large!

Red = If H_0 is true

Blue = If H_1 is true



Larger α increases power

$\alpha_{1\text{-tail}}$ = **.20**

β = .01

Power = .99

$\bar{X} - \mu = 3$ units

$\sigma_{\bar{X}} = 1$ unit

3 Common Types of Power

1. ***a priori* –** Before data, find *N* given:
   – α, $(1 - \beta)$, expected effect size

2. ***post hoc* –** After data, find power given:
   – α, *N*, **observed effect size**

3. ***Sensitivity* –** Before/after data, find detectable effect size given:
   – α, $(1 - \beta)$, *N*

Setting Power

- **Exp. 1**: "We sought to collect 80 participants… Sensitivity analysis indicated with power set at .80, we could detect an effect size as small as d_z = .317"

3. Sensitivity power b/c we don't know the size of the effect
    - α = .05
    - (1- β) = .80
    - N = 80

Setting Power

- **Exp. 2**: "Based on the observed effect size of d_z = .430 in Experiment 1, we sampled from 48 participants to set power at .80"

1. Use *a priori* power after getting an estimate of effect size in Exp. 1
    - α = .05
    - (1- β) = .80
    - Effect size = d_z = .430