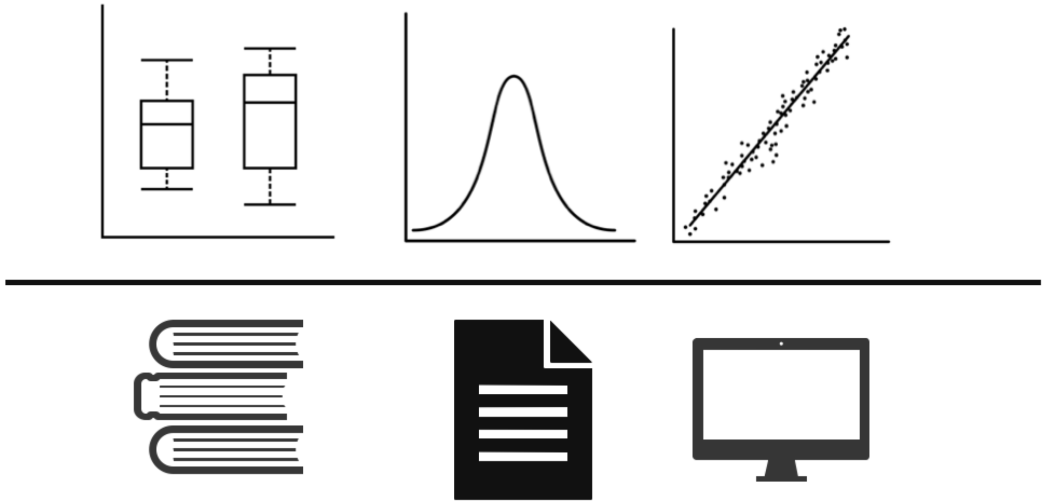
PSYC 2300

Introduction to Statistics



Lecture 07: Introduction to Statistical Significance

Outline for today

- Review parts of previous lectures
- Decisions in hypothesis testing
- Statistical power

Hypothesis

Hypothesis: a specific, clear, and testable proposition or predictive statement about the possible outcome of a scientific research study

When we use *samples* to approximate *populations*, however, we always have **sampling error** (difference between the sample statistic and the population parameter)

Sample Statistic $\overline{x}$

Population Parameter μ

Sampling Error ϵ

Statistical Hypotheses

Null Hypothesis: there is no change, no difference, or no relationship between variables

The difference observed is due only to *sampling error*, and not to any real effects in the population

Null Hypothesis

$$H_0 = \epsilon$$

Statistical Hypotheses

Research Hypothesis: there is change, a difference, or a relationship between variables

The difference observed is due to both sampling error and a real effect

Research Hypothesis

$$H_1 = \tau + \epsilon$$

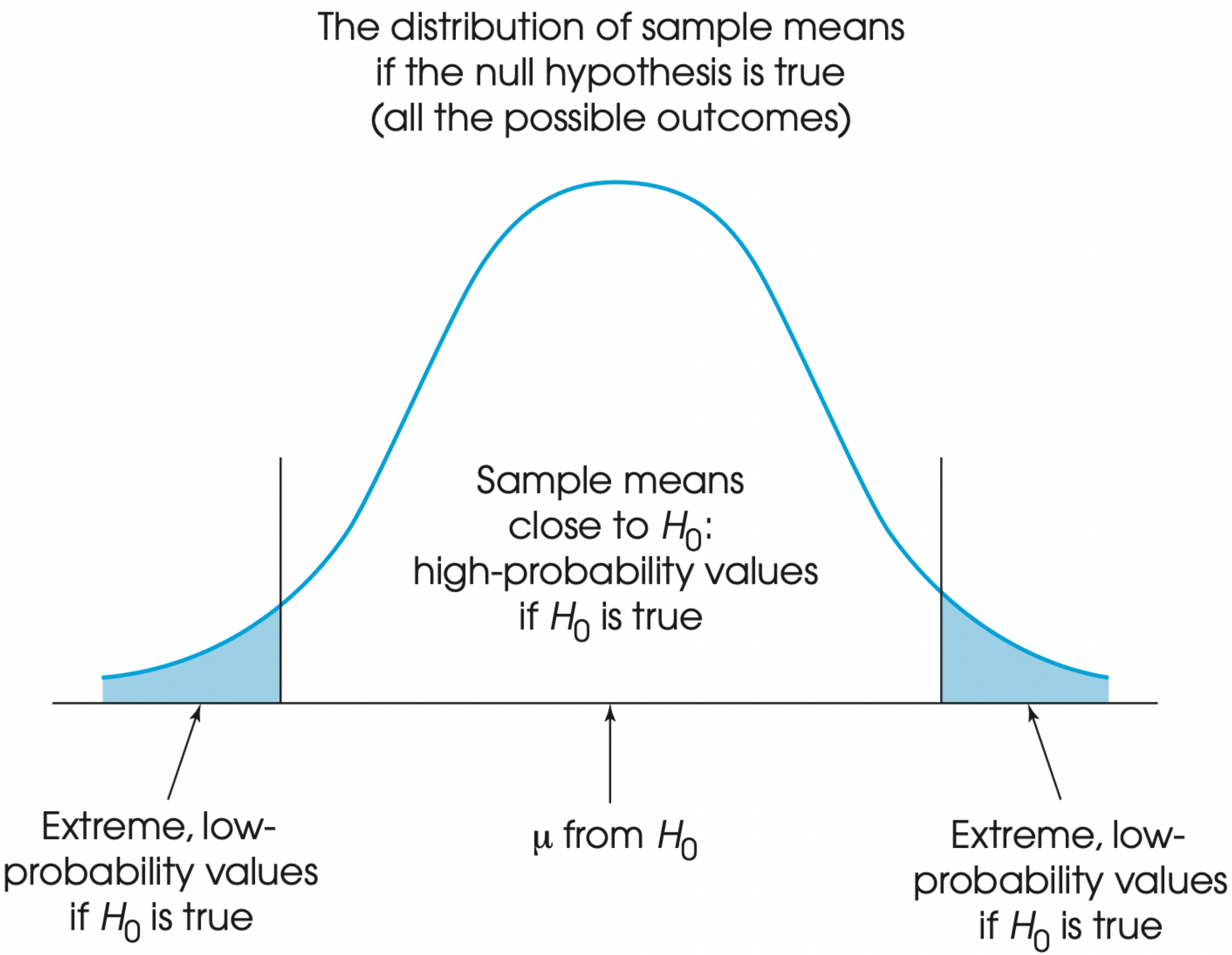
Standard of Evidence

Alpha: the probability value that is used to define which sample outcomes are considered very unlikely if H_0 is true

The most common $\alpha = .05$

Critical Region: The region (of the sampling distribution) that contains the sample outcomes that are considered *very unlikely* if H_0 is true

Critical Region



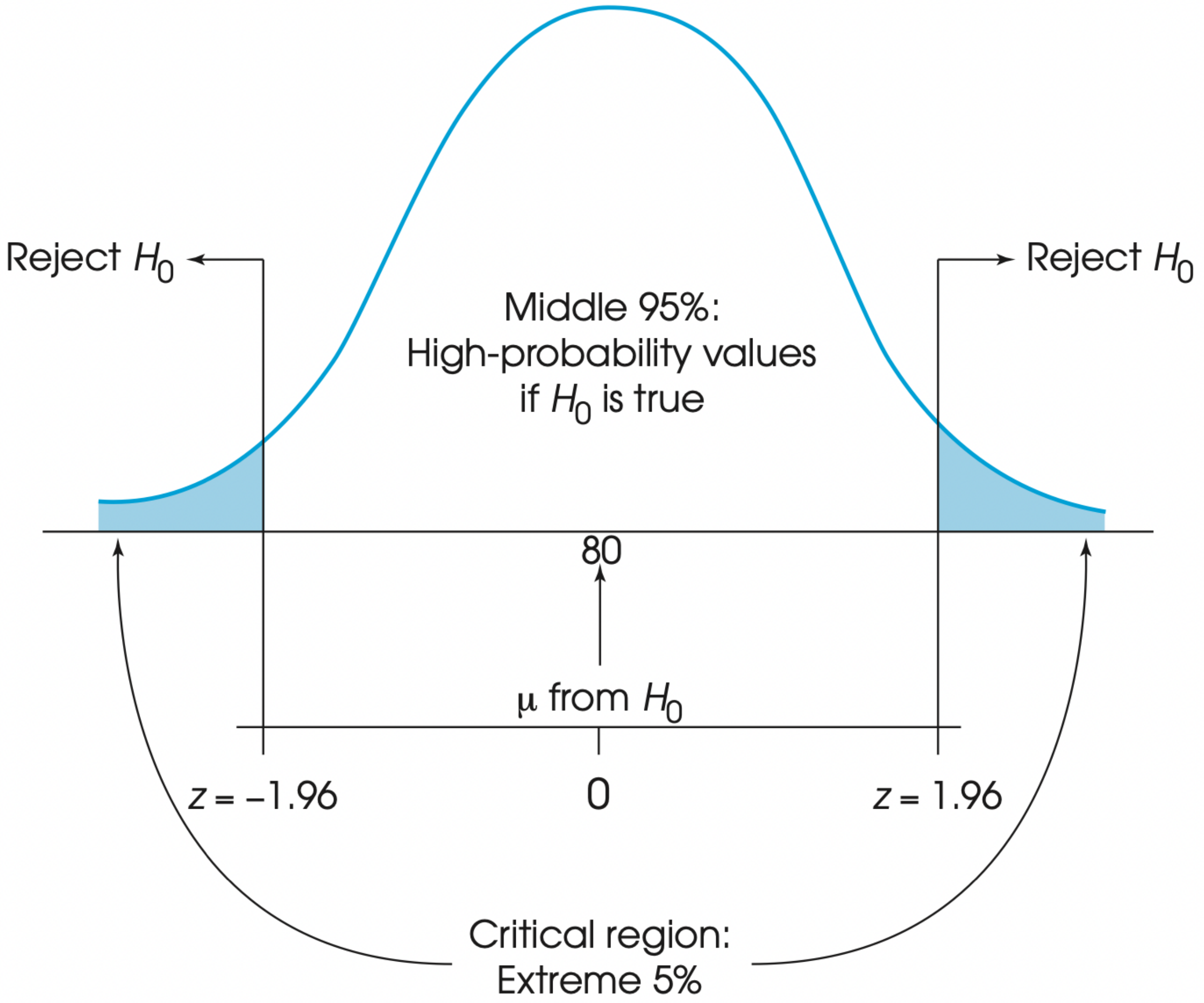
Compute a Test Statistic

Test statistic: A numerical summary of the degree to which a sample is unlike the samples predicted by the null hypothesis, H_0

***z*-test statistic**

$$z_{\overline{x}} = \frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}}$$

Standard of Evidence



p-value

***p*-value**: the probability of getting the observed or more extreme data, *assuming* the null hypothesis is true

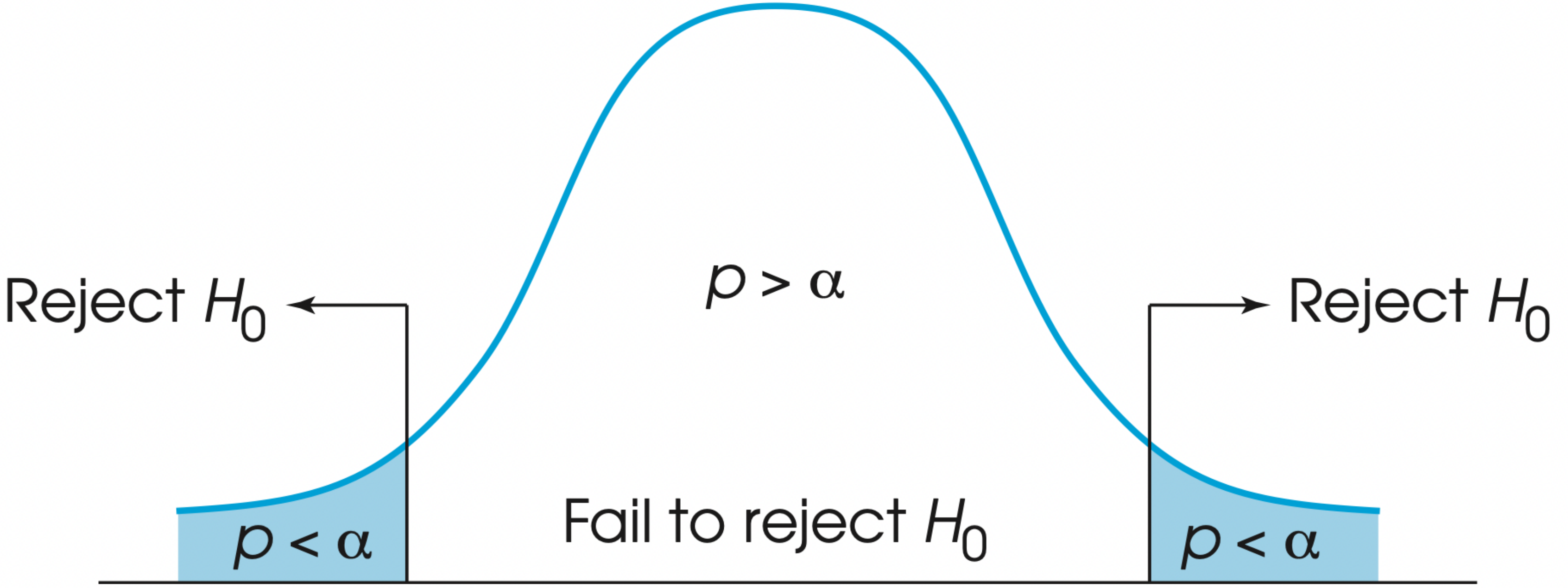
If $p < \alpha$

- The data you observe is **not** likely due to just sampling error
- Reject the null hypothesis

If $p > \alpha$

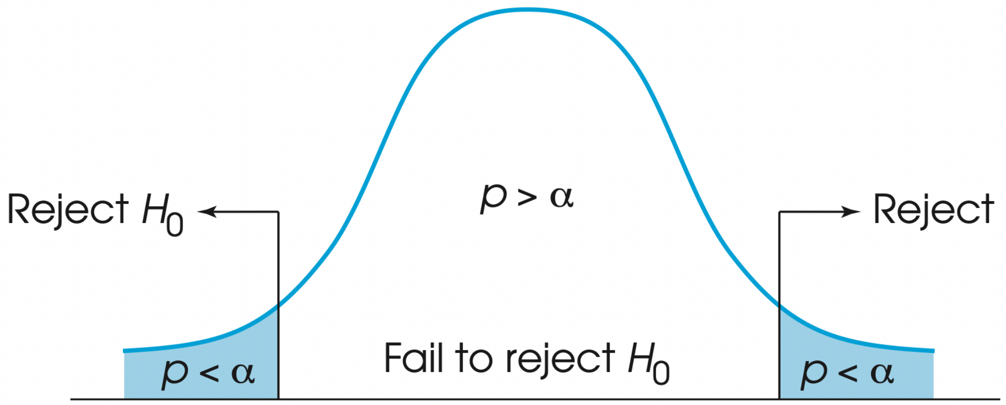
- The data is likely due to sampling error
- We fail to to reject the null hypothesis

p-value



Statistical Significance

Statistical Significance: A result is said to be *significant*, or *statistically significant*, if it is very unlikely to occur when the null hypothesis is true. That is, the result is sufficient to reject the null hypothesis. Thus, a treatment has a significant effect if the decision from the hypothesis test is to reject H_0



Statistical Hypotheses

Null Hypothesis

H_0: No real effect exists

$$\epsilon$$

Research Hypothesis

H_1: A real effect does exist

$$\tau + \epsilon$$

Because inferential statistics deals in *probabilities*, we always run the risk of making an error in our conclusions about the null and alternative hypotheses

Decisions in Hypothesis Testing

Possible Truths in the World

Truth

There is no effect, H_0	There is an effect, H_1

Because these are mutually exclusive and exhaustive, one of these has to be true and only one of them can be true

But *sampling error* is always acting, so occasionally we'll get samples that are different from the truth in the world, which will lead to erroneous conclusions

Possible Decision Outcomes

Truth

There is no true effect
H_0

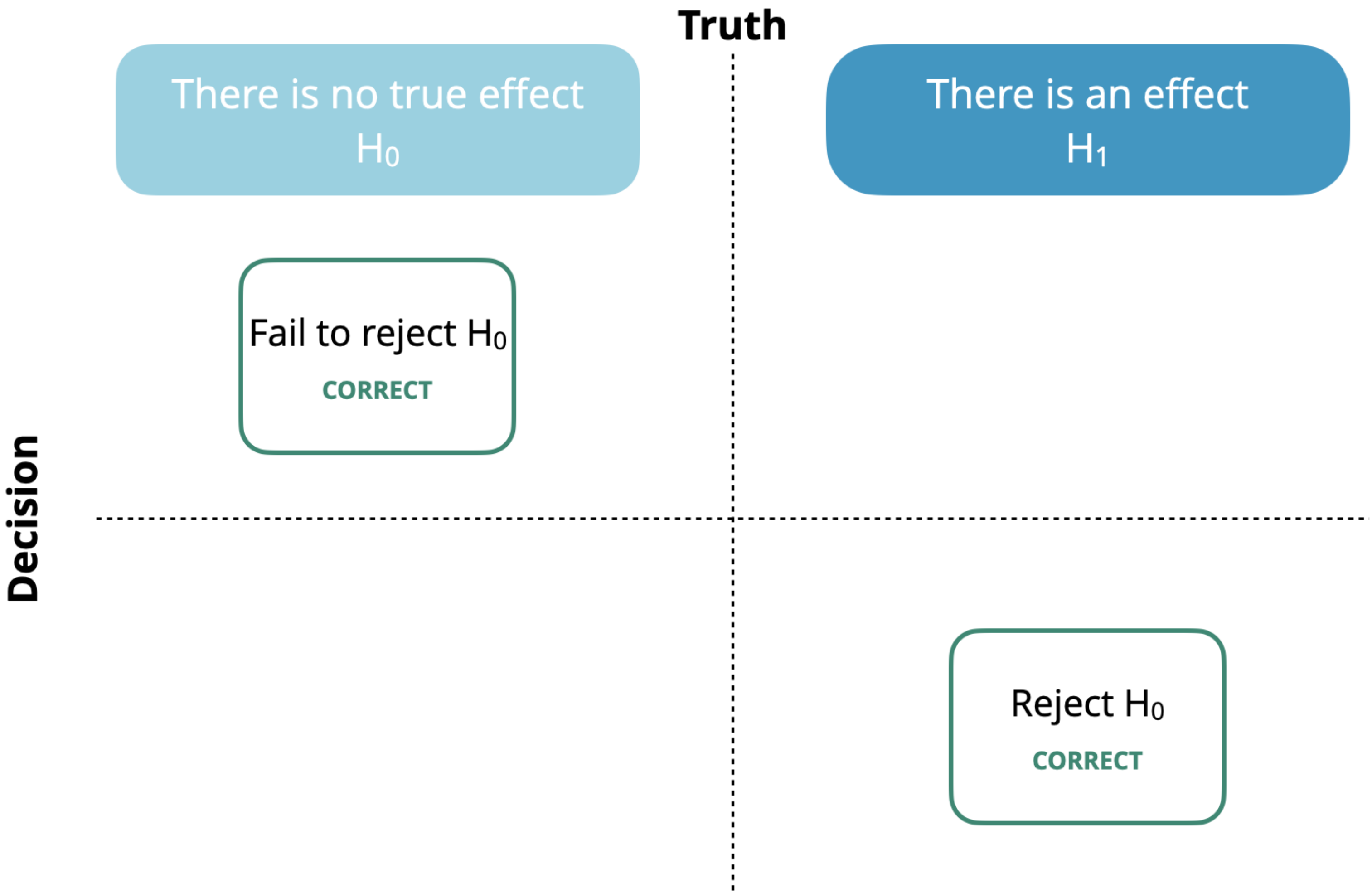
There is an effect
H_1

Possible Decision Outcomes

Truth

	There is no true effect H_0	There is an effect H_1
Decision	Fail to reject H_0 CORRECT	

Possible Decision Outcomes

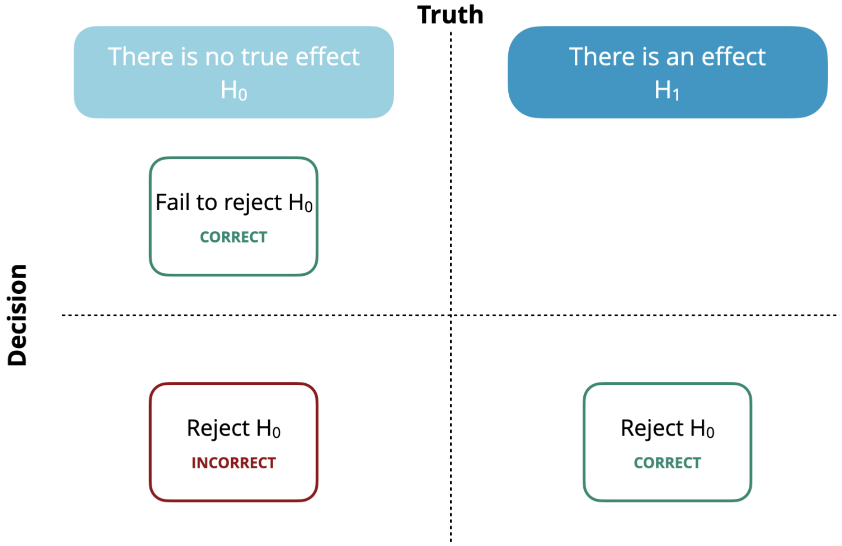


Possible Decision Outcomes

Decision \ Truth	There is no true effect H_0	There is an effect H_1
	Fail to reject H_0 **CORRECT**	
	Reject H_0 **INCORRECT**	Reject H_0 **CORRECT**

Type I Error

Type I error: Occurs when there is *no* effect present but the researcher rejects the null hypothesis



Also known as a '*false alarm*', '*Alpha error*', or '***False positive***'

Simply put: Saying there is an effect, when there is actually no effect

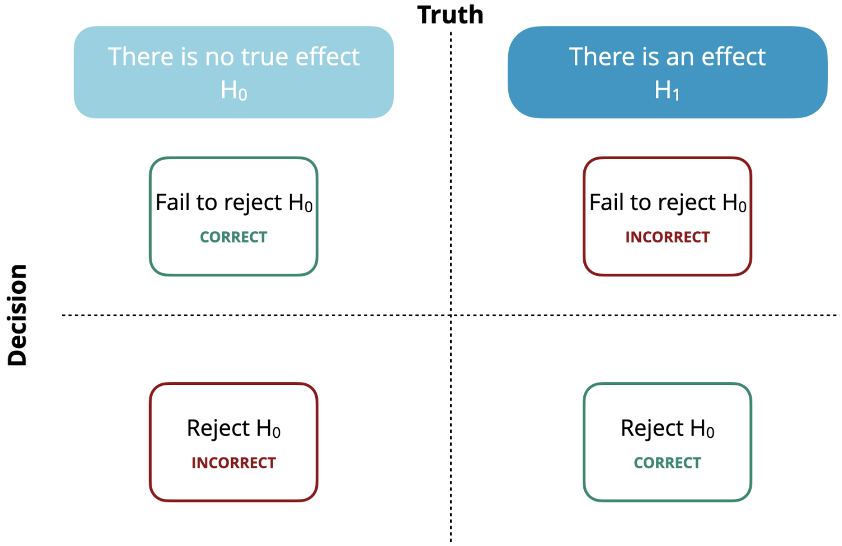
Decisions

Truth

Decision	There is no true effect H_0	There is an effect H_1
	Fail to reject H_0 **CORRECT**	Fail to reject H_0 **INCORRECT**
	Reject H_0 **INCORRECT**	Reject H_0 **CORRECT**

Type II Error

Type II error: Occurs when a real effect *is* present, but the researcher fails to reject the null hypothesis



Also known as a '*miss*' or '*Beta error*', or '***False negative***'

Simply put: Saying there is no effect, when there is actually an effect

Errors in Practical Terms

Errors in Practical Terms

Truth

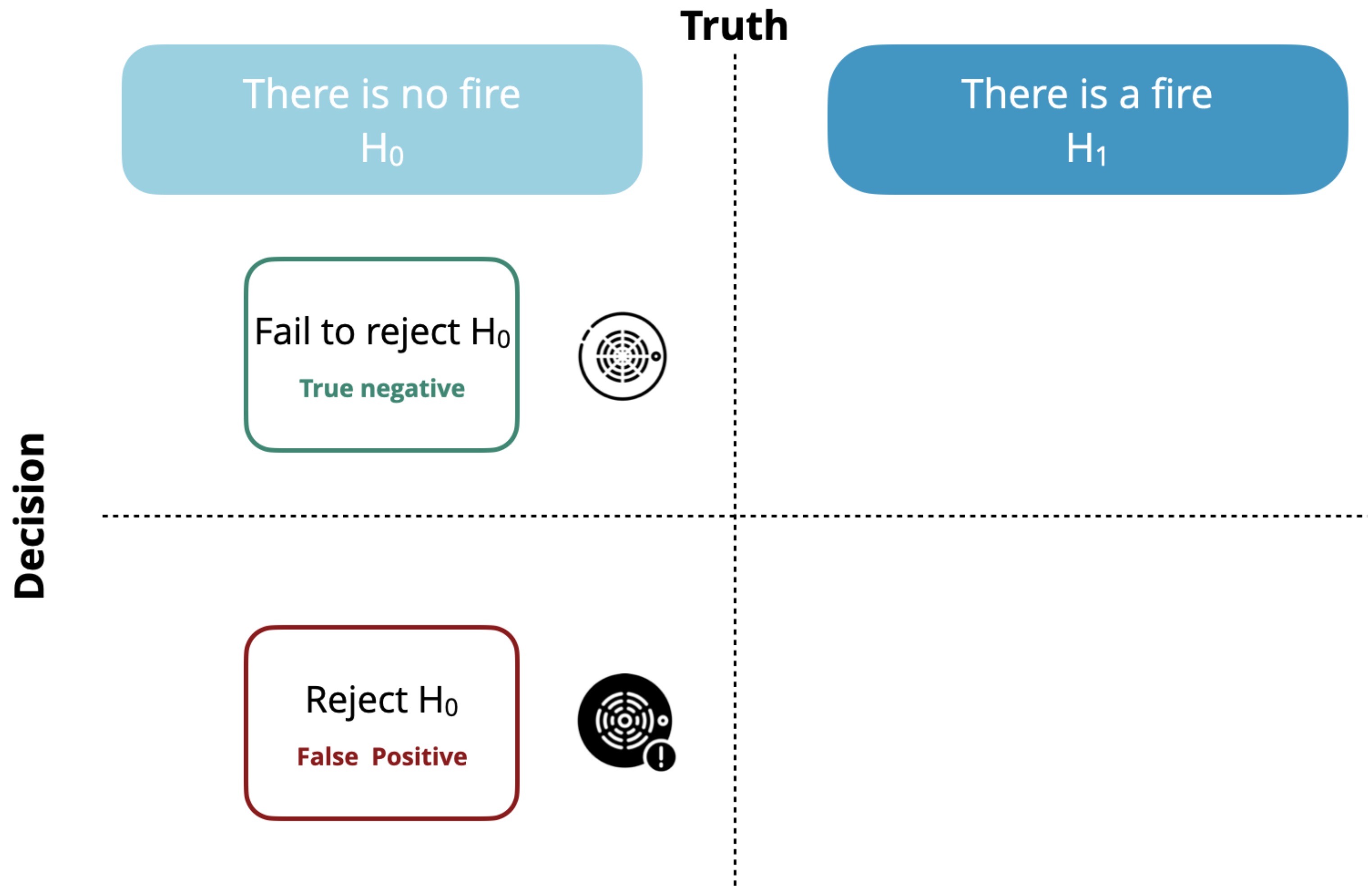
	There is no fire H_0	There is a fire H_1
Decision		

Errors in Practical Terms

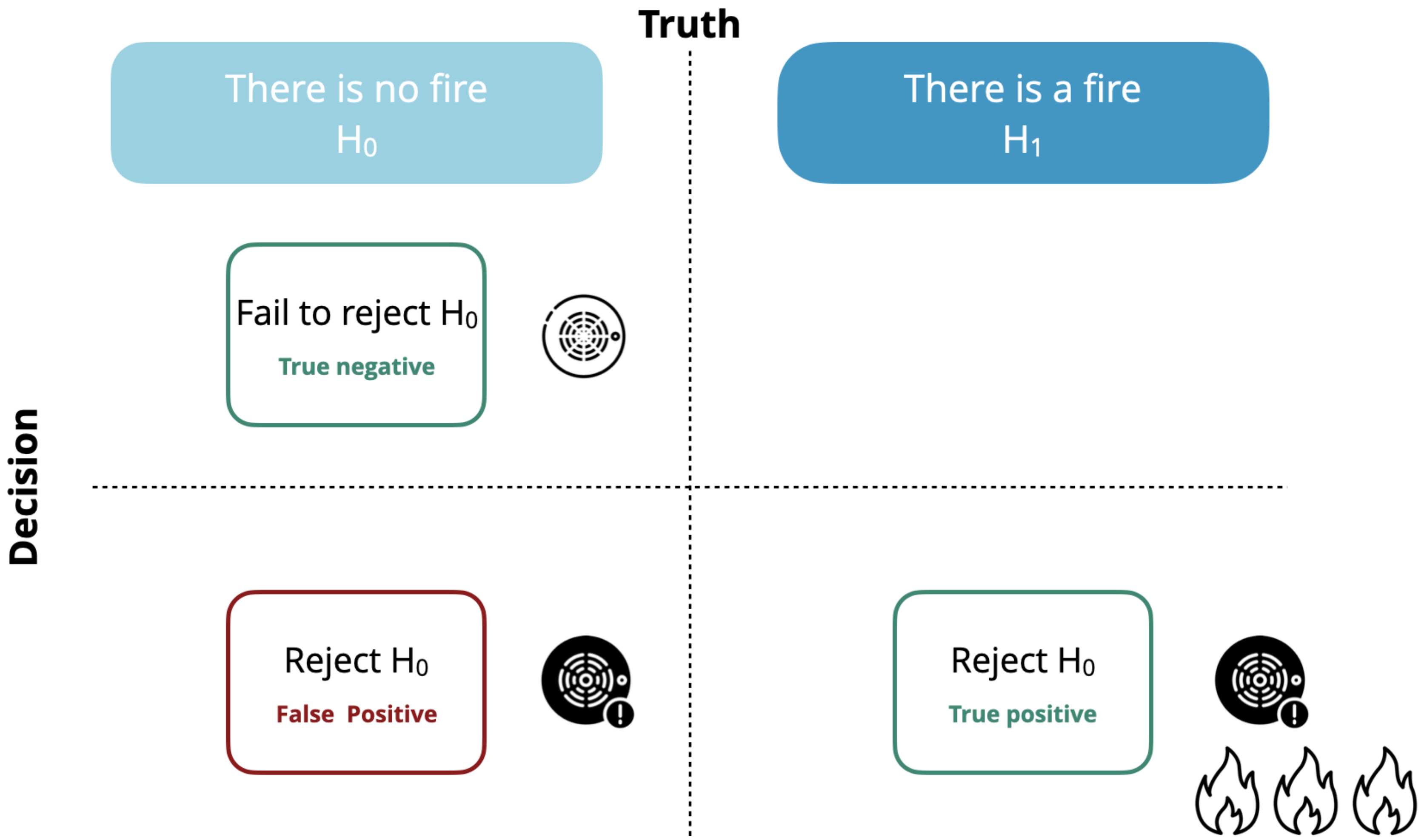
Truth

	There is no fire H_0	There is a fire H_1
Decision	Fail to reject H_0 **True negative** 	

Errors in Practical Terms



Errors in Practical Terms



Errors in Practical Terms

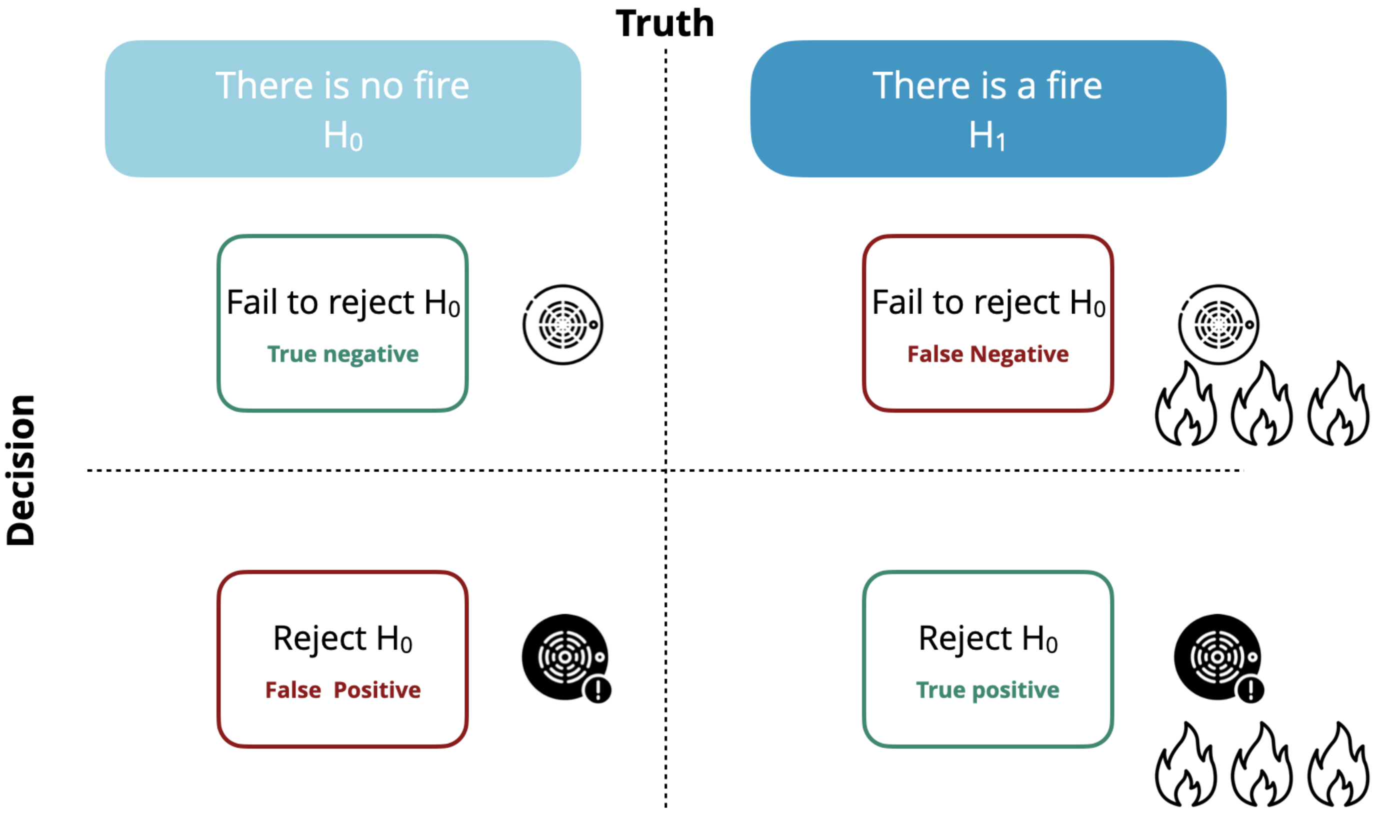
Decision \ Truth	There is no fire H_0	There is a fire H_1
	Fail to reject H_0 **True negative**	Fail to reject H_0 **False Negative**
	Reject H_0 **False Positive**	Reject H_0 **True positive**

Class Activity

Type I and Type II Errors

Class Activity

With your classmates, come up with an additional example of Type I and Type II errors. Draw them like the figure below.

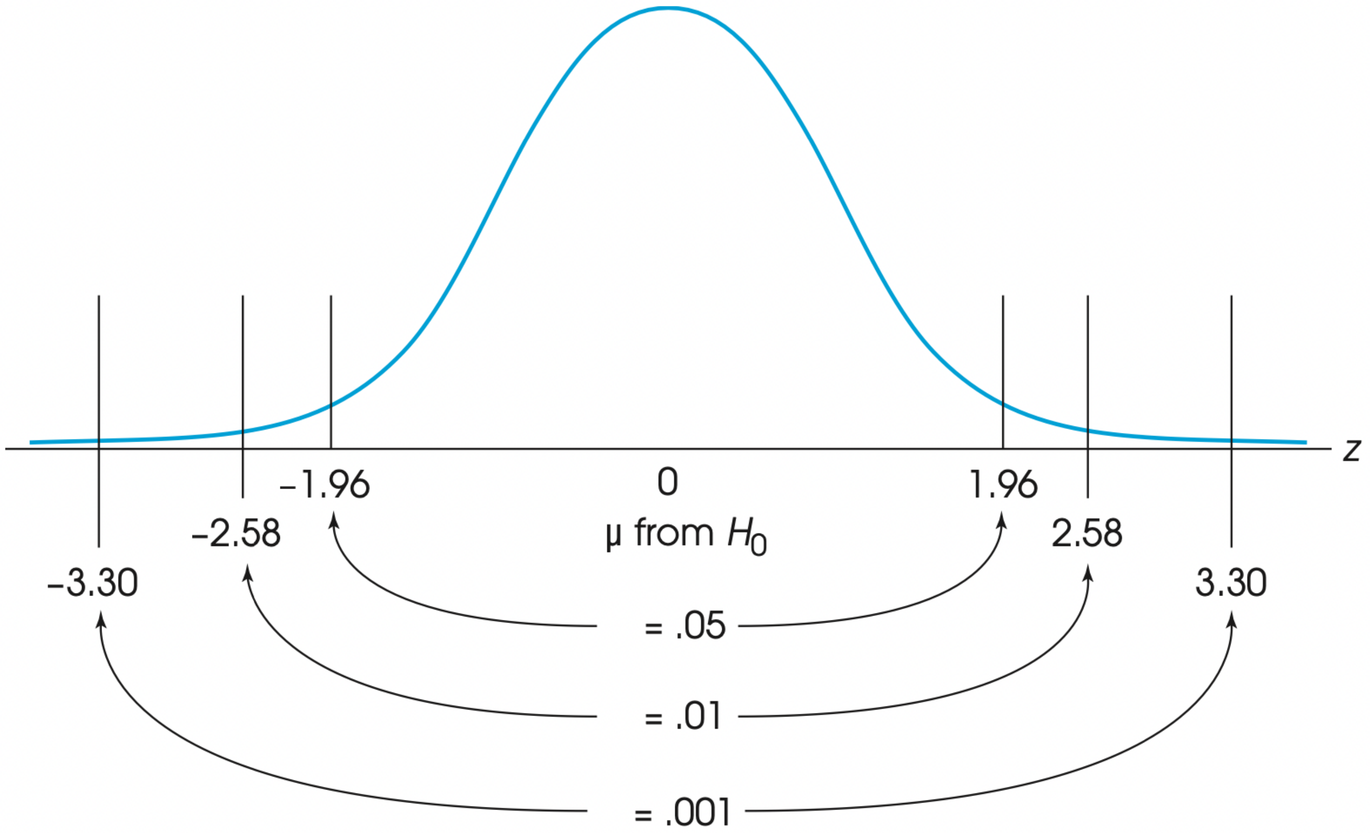


Influences of Type I Errors

- A Type I error occurs when we reject H_0 even though H_0 is true
- In other words, we found something in the critical region and called it "extreme enough" to be a real effect, when in reality, it wasn't
- Recall that boundaries of the critical region are determined by alpha, which is under our control

Critical Regions

The locations of the critical region boundaries for three different levels of significance:
$\alpha = .05, \alpha = .01, \alpha = .001$



Probability of a Type I Error

So, Type I error rates are under our control and their probability is simply equal to alpha (assuming that the true state of the world is that H_0 is true)

If H_1 is true, what proportion of the time could we have a false positive?

$$p(Type\ I\ error \mid H_1\ is\ True) = 0$$

$$p(Type\ I\ error \mid H_0\ is\ True) = \alpha$$

In other words:

- If H_1 is true, there is no risk of a Type I error
- If H_0 is true, the risk of a Type I error is α

Probability of a Type I Error

Truth

Decision	There is no true effect H_0		There is an effect H_1
	Fail to reject H_0 **SPECIFICITY**		Fail to reject H_0 **TYPE II ERROR**
	Reject H_0 **TYPE I ERROR**	α	Reject H_0 **POWER**

Probability of a Type II Error

$$p(Type\ II\ error \mid H_1\ is\ True) = \beta$$

$$p(Type\ II\ error \mid H_0\ is\ True) = 0$$

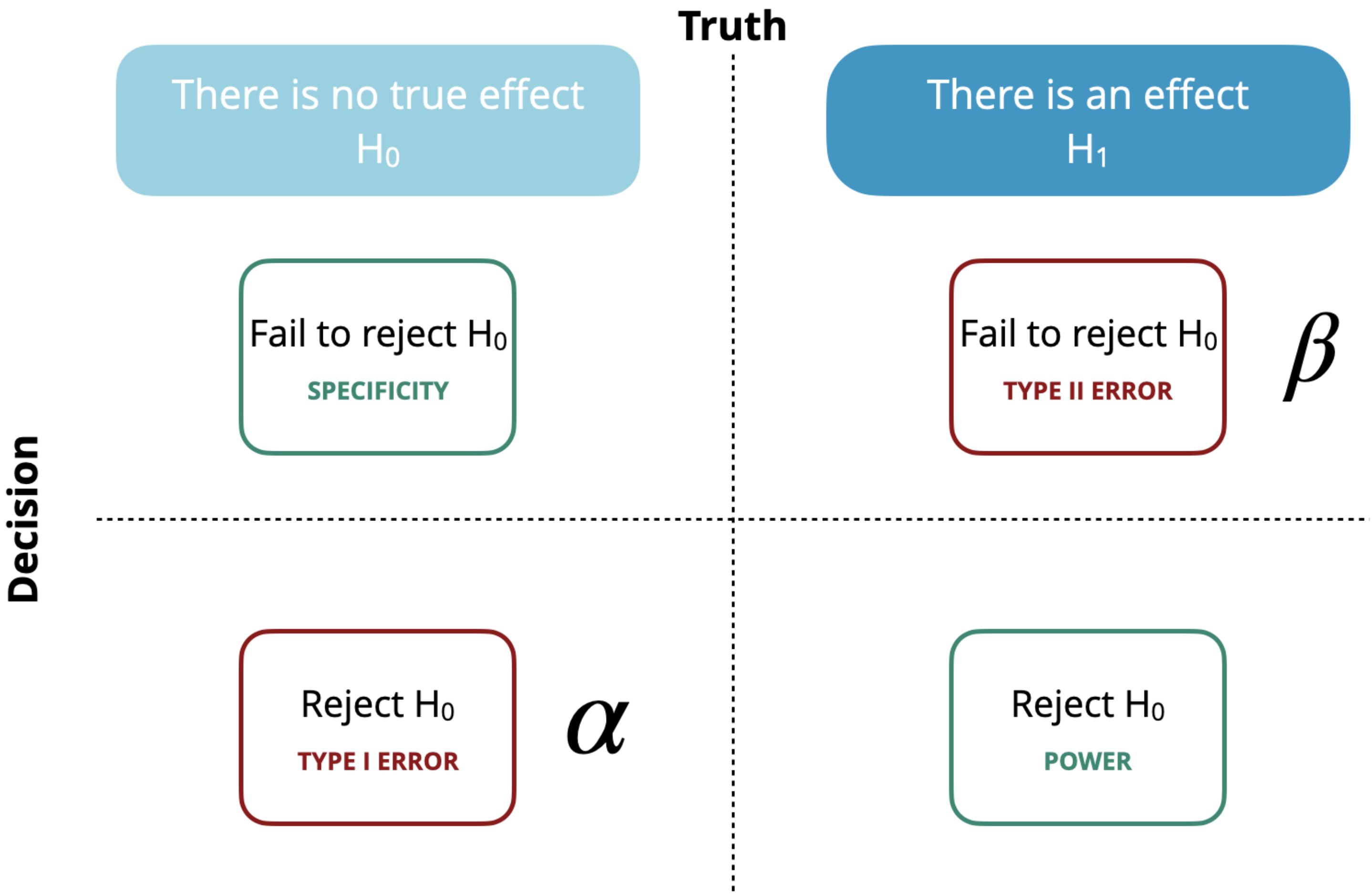
In other words:

- If H_1 is true, the risk of a Type II error is β
- If H_0 is true, there is no risk of a Type II error

Beta

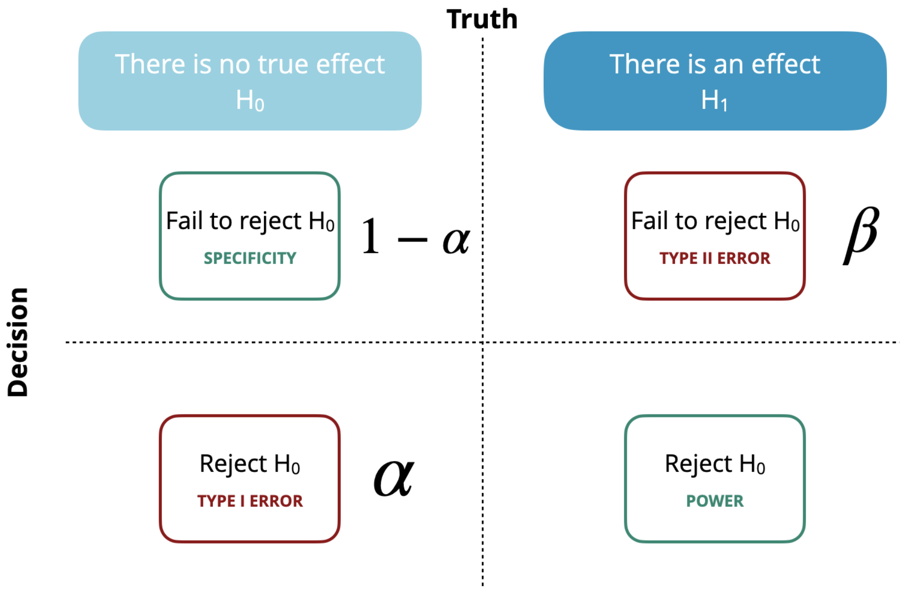
β

Probability of a Type II Error

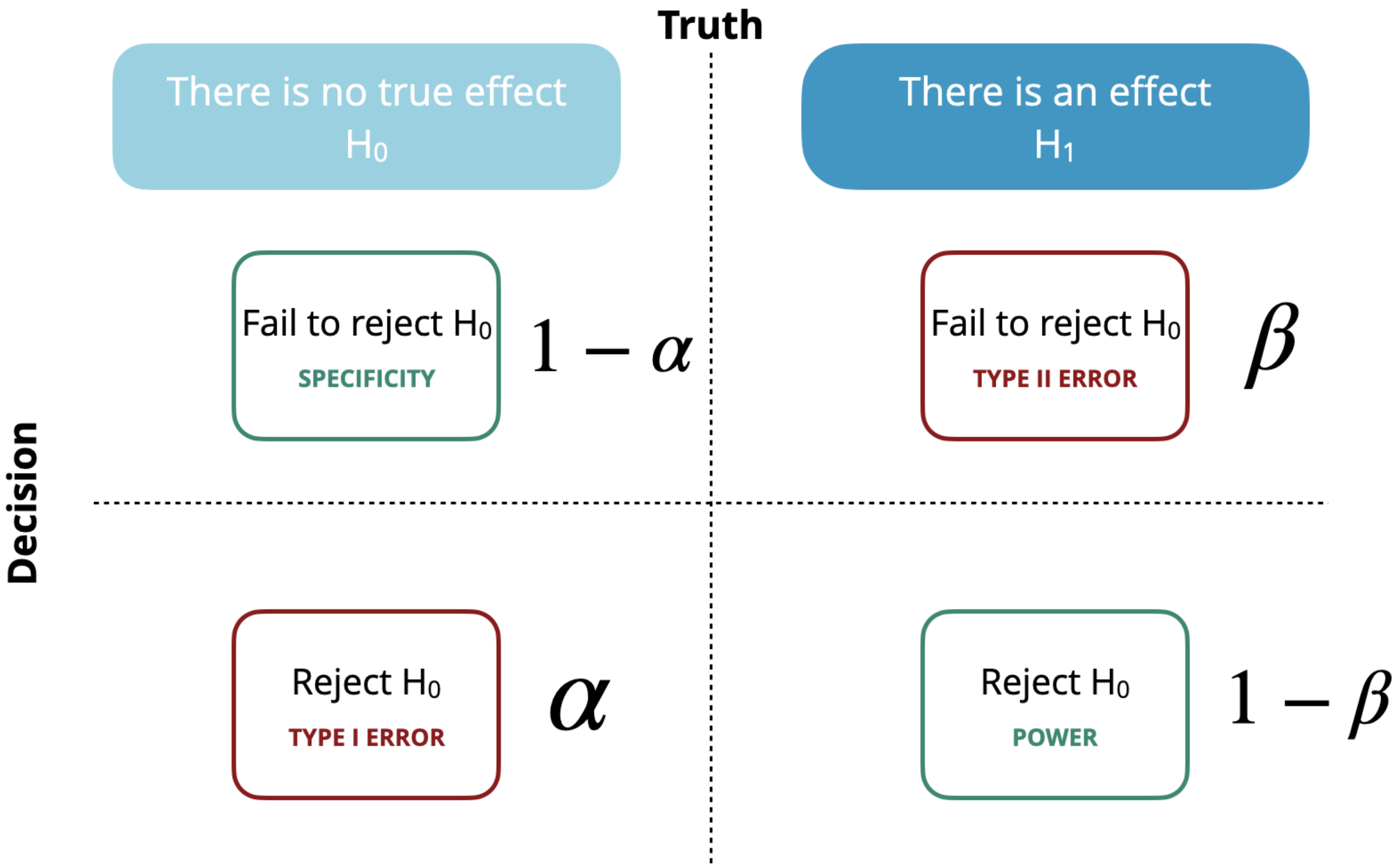


Probability of Specificity

Specificity: The specificity of a test (also called the True Negative Rate) is the probability of failing to reject H_0 when H_0 is true



Statistical Power



Alpha and Beta

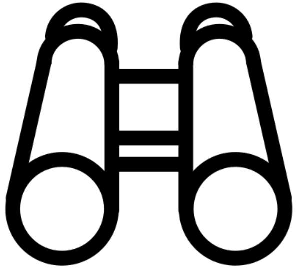
Alpha: probability of a false positive (Type I error)

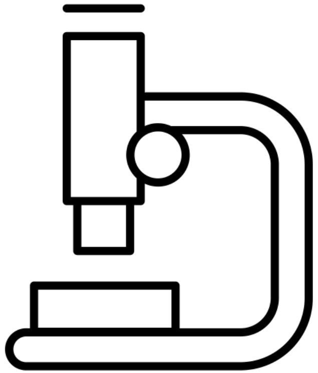
Beta: probability of a false negative (Type II error)

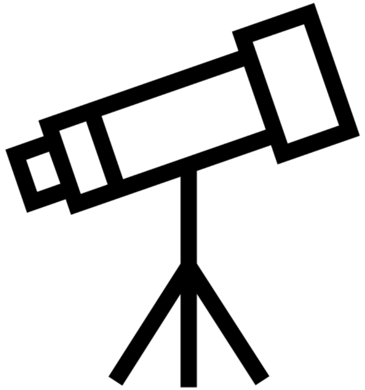
- We know alpha ahead of time and it is under our control, but what about beta?
- We often don't know what it is equal to ahead of time, but beta comes from statistical power, $1 - \beta$, which we will discuss next

Statistical Power

Statistical Power







Statistical Power

Statistical power: the probability that the test will correctly reject a false null hypothesis. That is, power is the probability that the test will identify an effect if one really exists

Simply put: how well a test can *detect* a real effect and reject the null hypothesis when the null hypothesis is, in fact, false

Statistical Power

In the population

$\mu = 100$

$\sigma = 15$

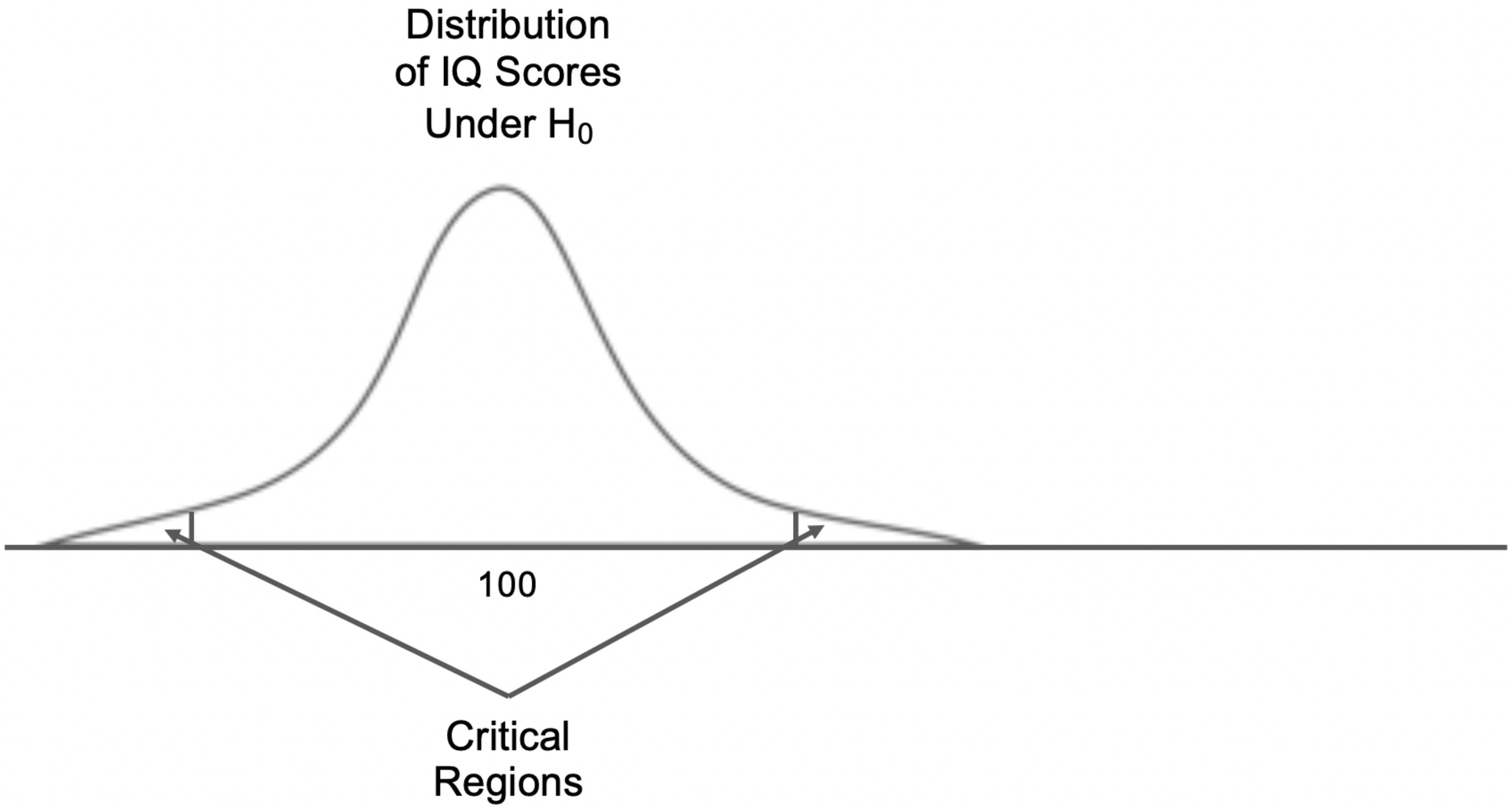
Our sample

$$n = 100$$

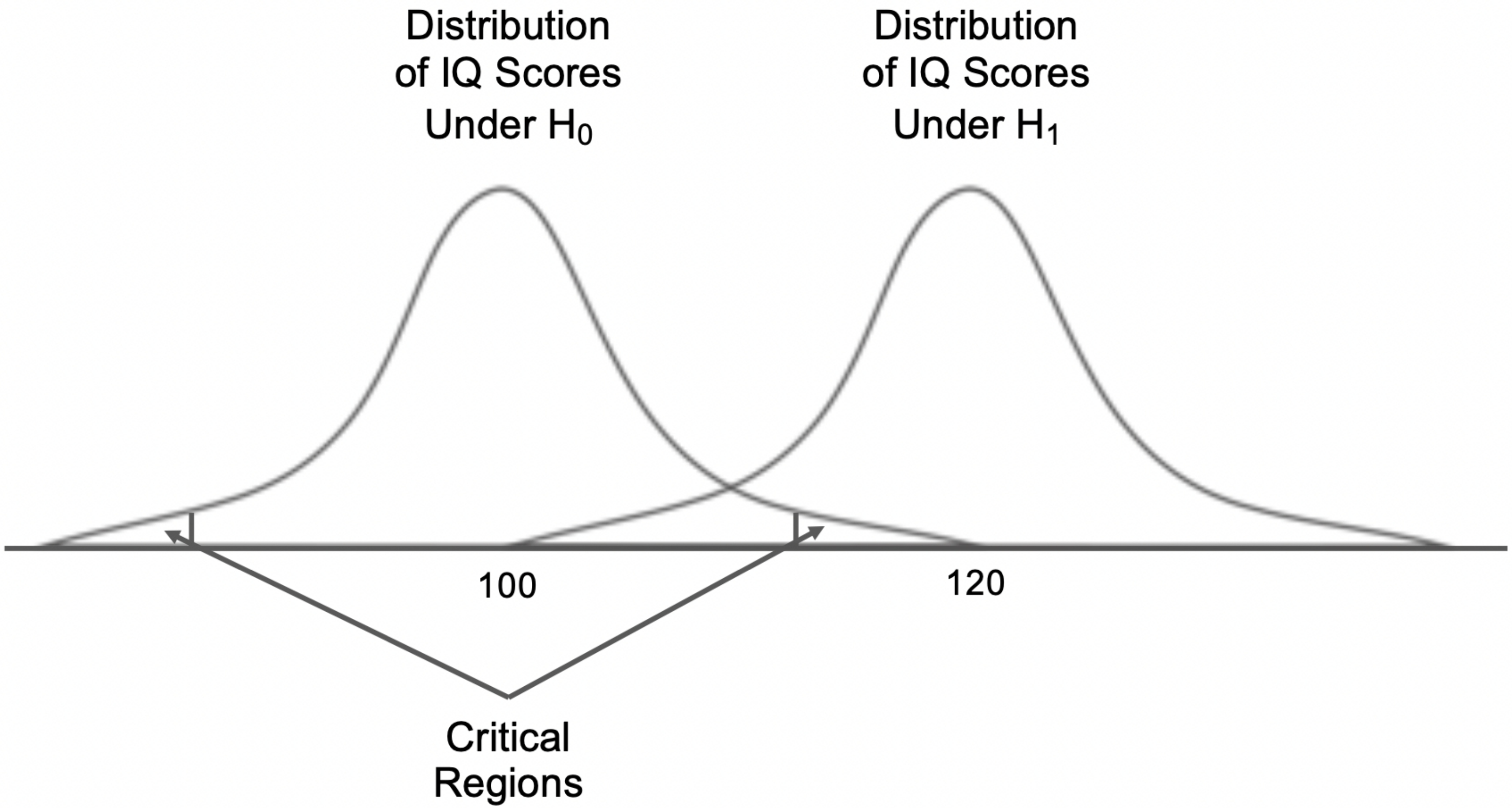
$$\overline{X}_{IQ} = 120$$

$$s_{IQ} = 15$$

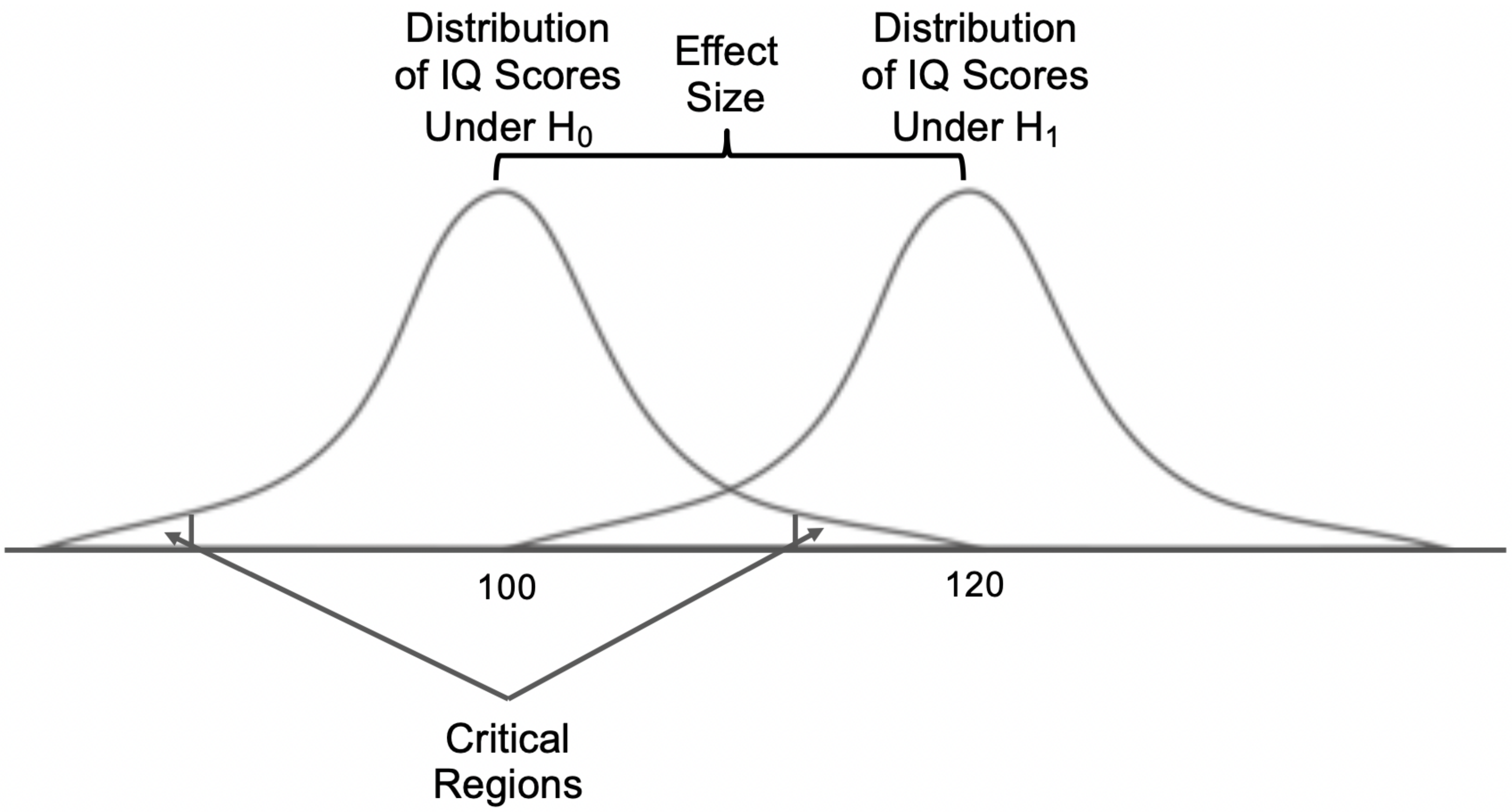
Statistical Power



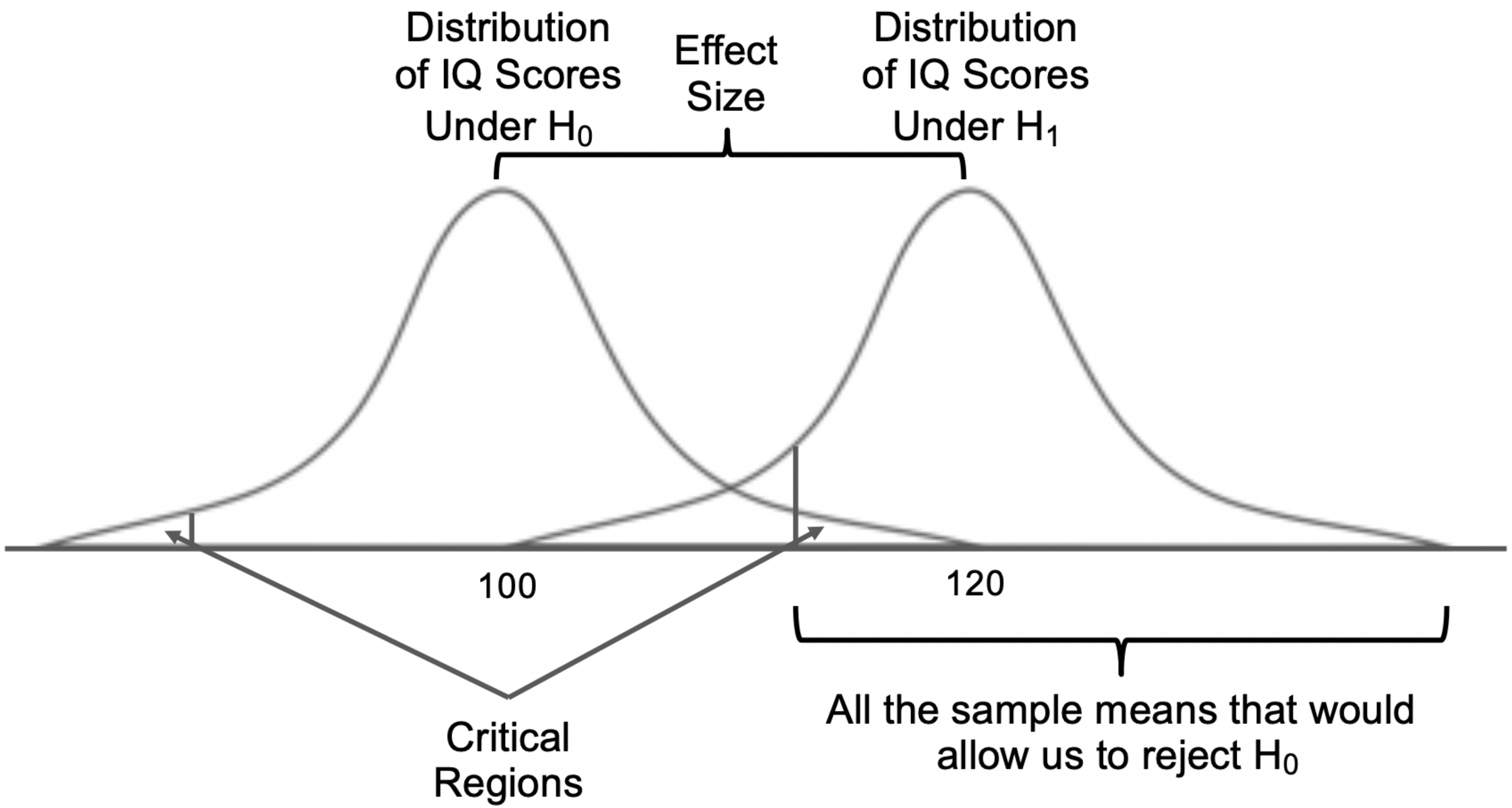
Statistical Power



Statistical Power



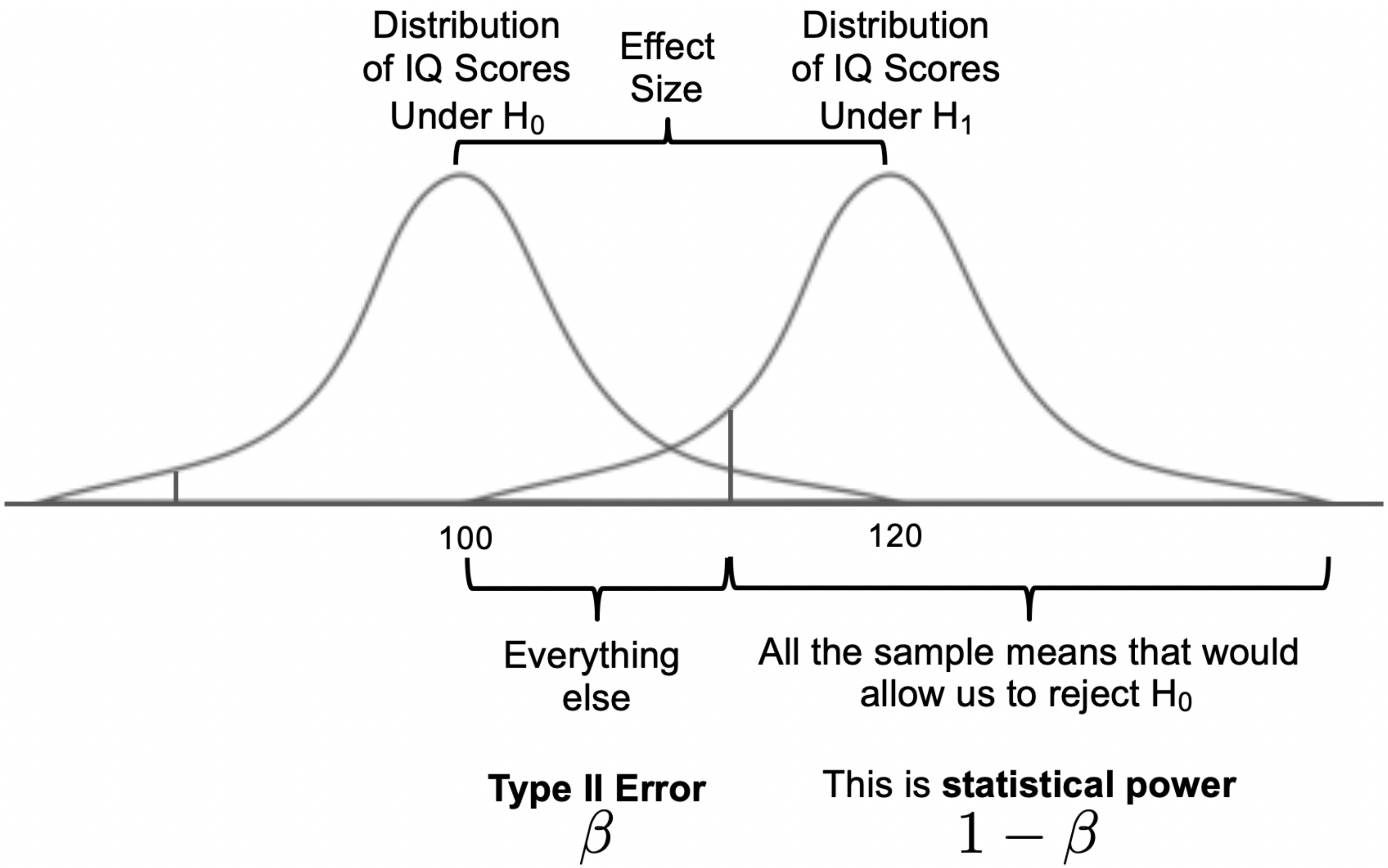
Statistical Power



This is **statistical power**

$$1 - \beta$$

Statistical Power



Influences of Statistical Power

- The **size of an effect** in the population
  - Bigger effect size → more statistical power

Effect Size

Cohen's *d* for one-sample *z*-test

$$d_z = \frac{\bar{x} - \mu}{\sigma}$$

The size of an effect here would be how large our numerator is

$\bar{x}$ estimates the mean under H_1

μ is the mean under H_0

σ is the standard deviation under H_0

Influences of Statistical Power

- The **size of an effect** in the population
  - Bigger effect size → more statistical power
- **Variability** in the population
  - Less variability → more statistical power

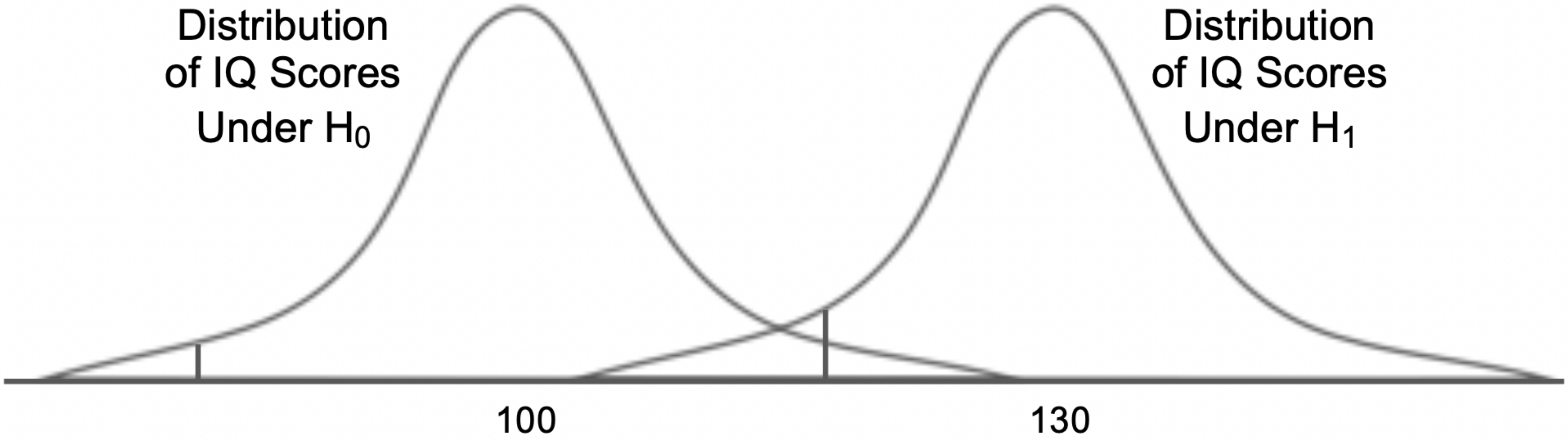
Variability

z Test Statistic

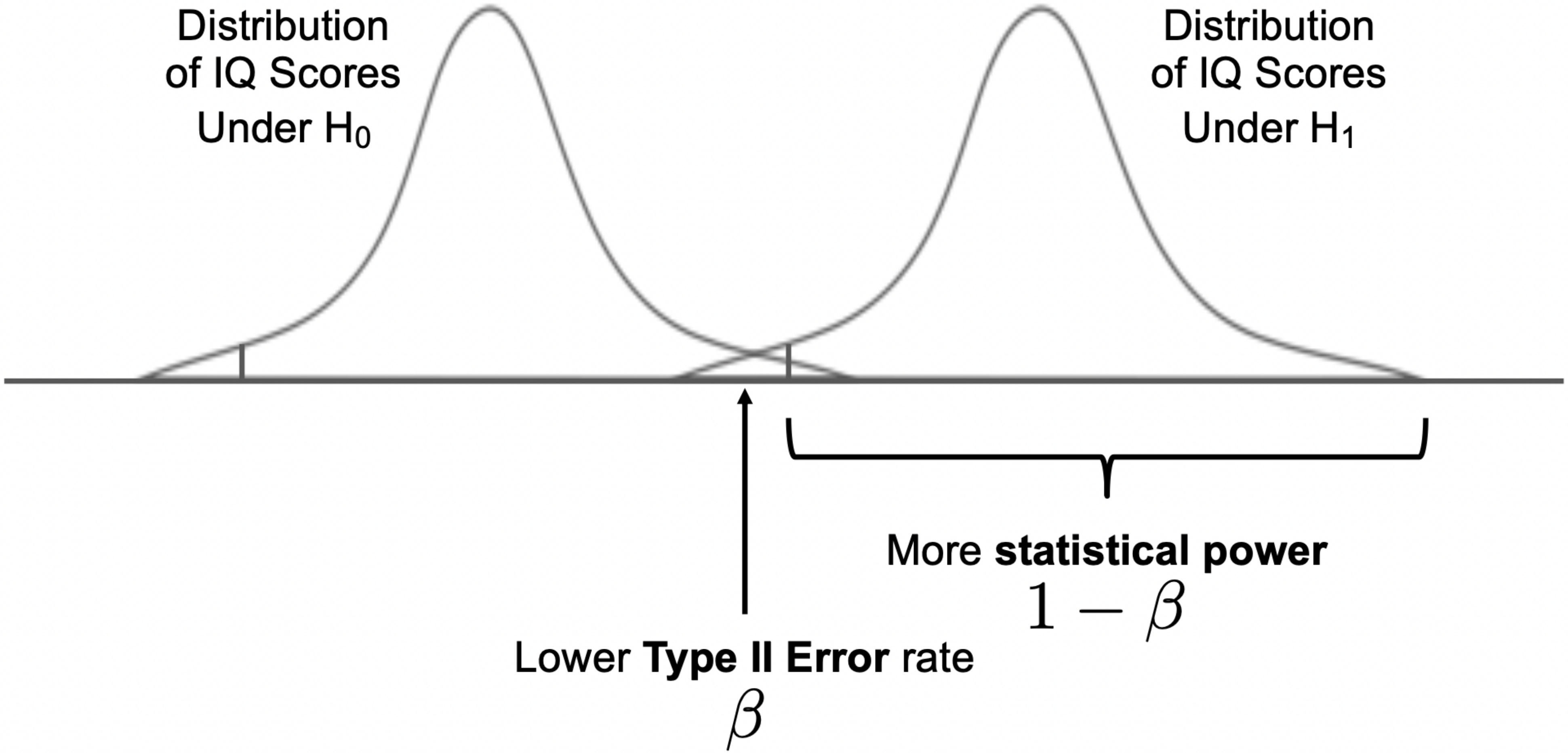
$$z_{\overline{x}} = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}}$$

The more variability in the population, the larger the standard error, the smaller the test statistic

Variability



Variability



Influences of Statistical Power

- The **size of an effect** in the population
  - Bigger effect size → more statistical power
- **Variability** in the population
  - Less variability → more statistical power
- **Sample size**
  - Larger sample size → more statistical power

Sample Size

As sample size increases, standard error decreases, the test statistic increases, and statistical power increases

z Test Statistic

$$z_{\overline{x}} = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}}$$

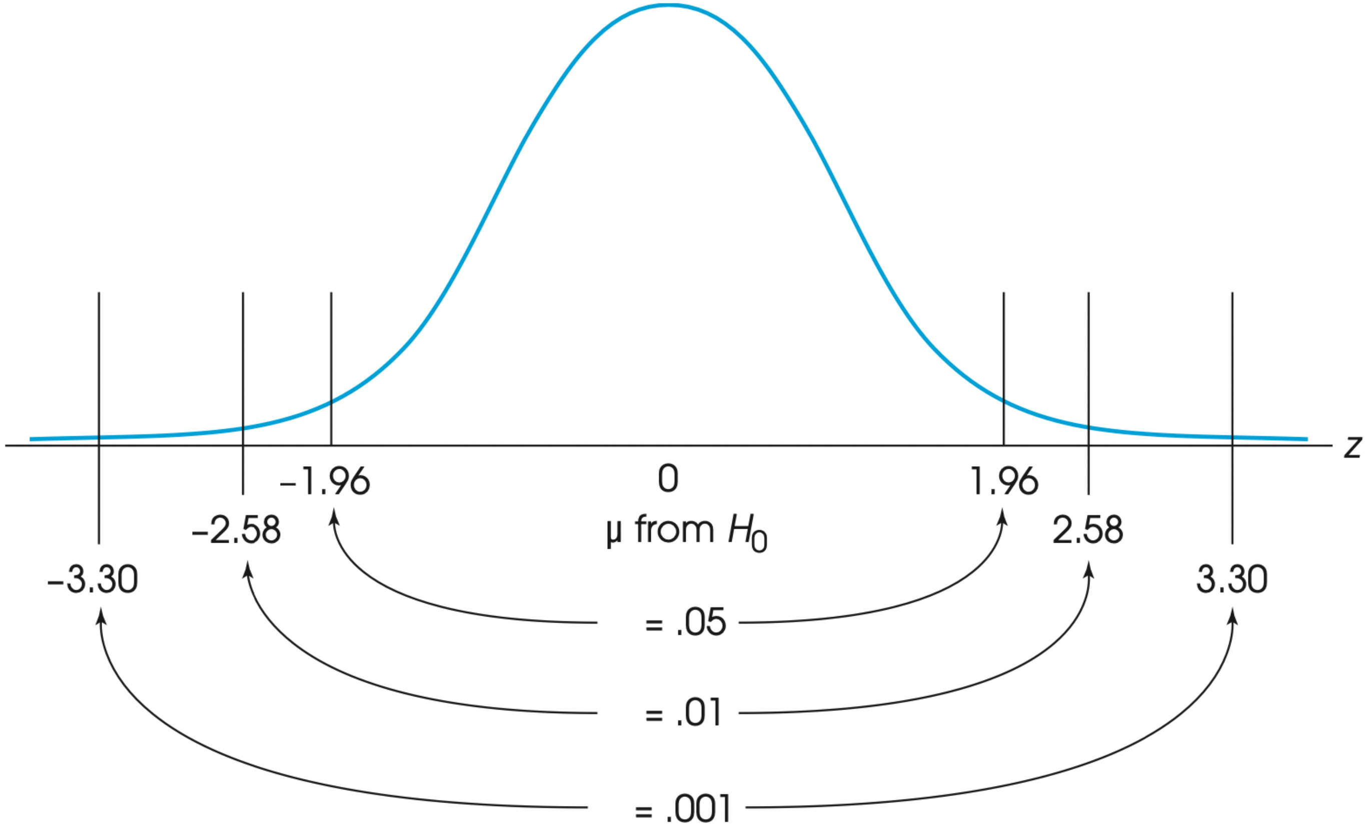
Standard Error

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Influences of Statistical Power

- The **size of an effect** in the population
  - Bigger effect size → more statistical power
- **Variability** in the population
  - Less variability → more statistical power
- **Sample size**
  - Larger sample size → more statistical power
- **Alpha level**: Bigger alpha level → more statistical power

Alpha Level



Which error rates (Type I, Type II) should you aim for?

Setting Alpha and Statistical Power

Usually we set $\alpha = 0.05$ and statistical power, $1 - \beta = 0.80$

Alpha

This means in the *long run*,
5% of our results will be Type I errors

5% false positive rate

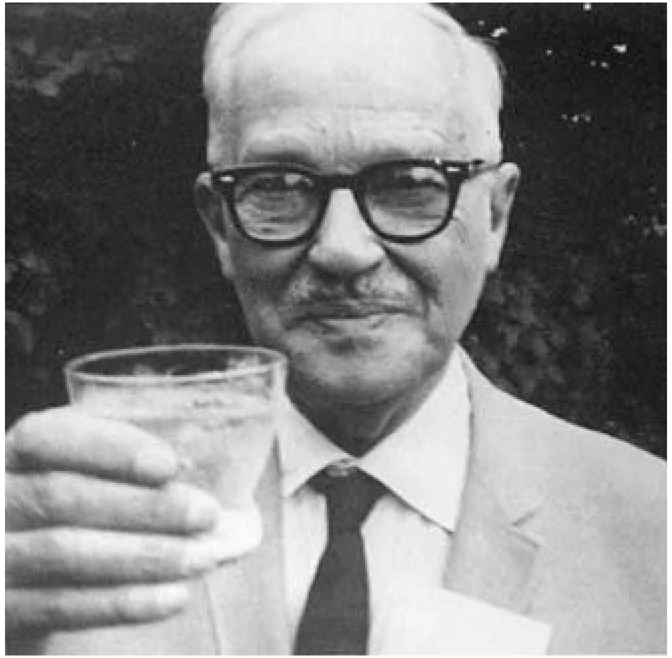
Statistical power

This means in the *long run*,
20% of our results will be Type II errors

20% false negative rate

Setting Alpha and Statistical Power

> Is it more serious to convict an innocent man, or to acquit a guilty?
>
> Determining how the balance must be struck should be left to the investigator.
>
> — *Neyman & Pearson, 1933*



Jerzy Neyman

Class Activity

Class Activity: Settting Error Rates

With your classmates, determine:

What is an acceptable *false positive rate* (Type I error), and why?

What is an acceptable *false negative rate* (Type II error), and why?

Example Research Questions

- We are studying a new education program for increasing reading comprehension, and the intervention will cost $300,000 to implement
- We are studying a new forensic interviewing technique to better differentiate liars from truth-tellers
- We are studying whether violent video games cause aggression
- We are studying whether social media use is associated with depression and anxiety symptoms

Next time

Lecture

- Midterm Review

Reading

- Chapter Nine

Quiz

- Quiz 2 is due tonight at 11:59pm MT
  - Lecture 4-6, Ch.6-8