HOW TO APPLY THE MULTIPHASE OPTIMIZATION STRATEGY (MOST) IN YOUR INTERVENTION DEVELOPMENT RESEARCH

Module 4
Some conceptual and technical aspects of the factorial experiment

Lesson 3: Understanding the basics of powering a factorial experiment

This course was developed by

Linda M. Collins

School of Global Public Health

New York University

(narrator)

Kate Guastaferro

College of Health and Human Development

The Pennsylvania State University

In the previous lesson you learned how to:

- Interpret main effects and interaction effects

In this lesson you will learn how to:

- Understand the basics of powering a factorial experiment

Review of the logic of hypothesis testing

- To analyze the data from a factorial optimization trial you will use factorial analysis of variance (ANOVA)
- We are going to approach this discussion using the *t*-test, which is closely related

Review of the logic of hypothesis testing

- Suppose you are interested in whether two means, μ_1 and μ_2, are different
- You cannot directly observe the parameters μ_1 and μ_2
- Instead, you can estimate them in a sample via the statistics $\bar{Y}_1$ and $\bar{Y}_2$

Review of the logic of hypothesis testing

- Null hypothesis: $H_0 \colon \mu_1 - \mu_2 = 0$
- Suppose H_0 is true. Would you expect to observe $\bar{Y}_1 - \bar{Y}_2 = 0$ every time?
- Suppose H_0 is false. Would you expect to observe $\bar{Y}_1 - \bar{Y}_2 \neq 0$ every time?
- But you need to draw a conclusion about whether the null hypothesis is false. What can you do?

Review of the logic of hypothesis testing

- Compute this test statistic:

$$t_{observed} = \frac{\bar{Y}_1 - \bar{Y}_2}{s_Y/n}$$

- The distribution of t when H_0 is true <u>is known</u>

Review of the logic of hypothesis testing

Because you know the t distribution under H_0 you can pose this question:

ASSUME FOR THE MOMENT THAT H_0 IS TRUE. WHAT IS THE PROBABILITY THAT I WOULD OBSERVE A t STATISTIC AS LEAST AS LARGE AS $t_{observed}$?

The t distribution provides this probability, which is often called the p-value.

Review of the logic of hypothesis testing

ASSUME FOR THE MOMENT THAT H_0 IS TRUE. WHAT IS THE PROBABILITY THAT I WOULD OBSERVE A t STATISTIC AS LEAST AS LARGE AS $t_{observed}$?

- If the answer is “the probability is low” decide “H_0 is false”
- If the answer is “the probability is not low” decide “cannot reject H_0”

This is hypothesis testing!

Review of the logic of hypothesis testing

- Your decision will be either correct or incorrect. There are two different types of errors that can be made.
- Type I error = rejecting H_0 when H_0 is true = mistakenly concluding an effect exists when it does not
- α = Type I error rate = probability of rejecting H_0, given that H_0 is true

Review of the logic of hypothesis testing

- Your decision will be either correct or incorrect. There are two different types of errors that can be made.
- Type II error = failing to reject H_0 when H_0 is false = <u>mistakenly</u> overlooking an effect
- β = Type II error rate = probability of failing to reject H_0, given that H_0 is false

Review of the logic of hypothesis testing

- β = Type II error rate = probability of failing to reject H_0, given that H_0 is false
- NOTE β is used in statistics to refer to two different things: the regression coefficient and the probability of a Type II error

Statistical power 101

- Note that both Type I and Type II error rates are conditional probabilities of making incorrect decisions
  - Type I error rate (α): *P*(rejecting $H_0|H_0$ true)
  - Type II error (β): *P*(not rejecting $H_0|H_0$ false)

- Power is also a conditional probability, but of making a correct decision when H_0 is false
  - Power = *P*(rejecting $H_0|H_0$ false) = $1 - \beta$

Statistical power 101

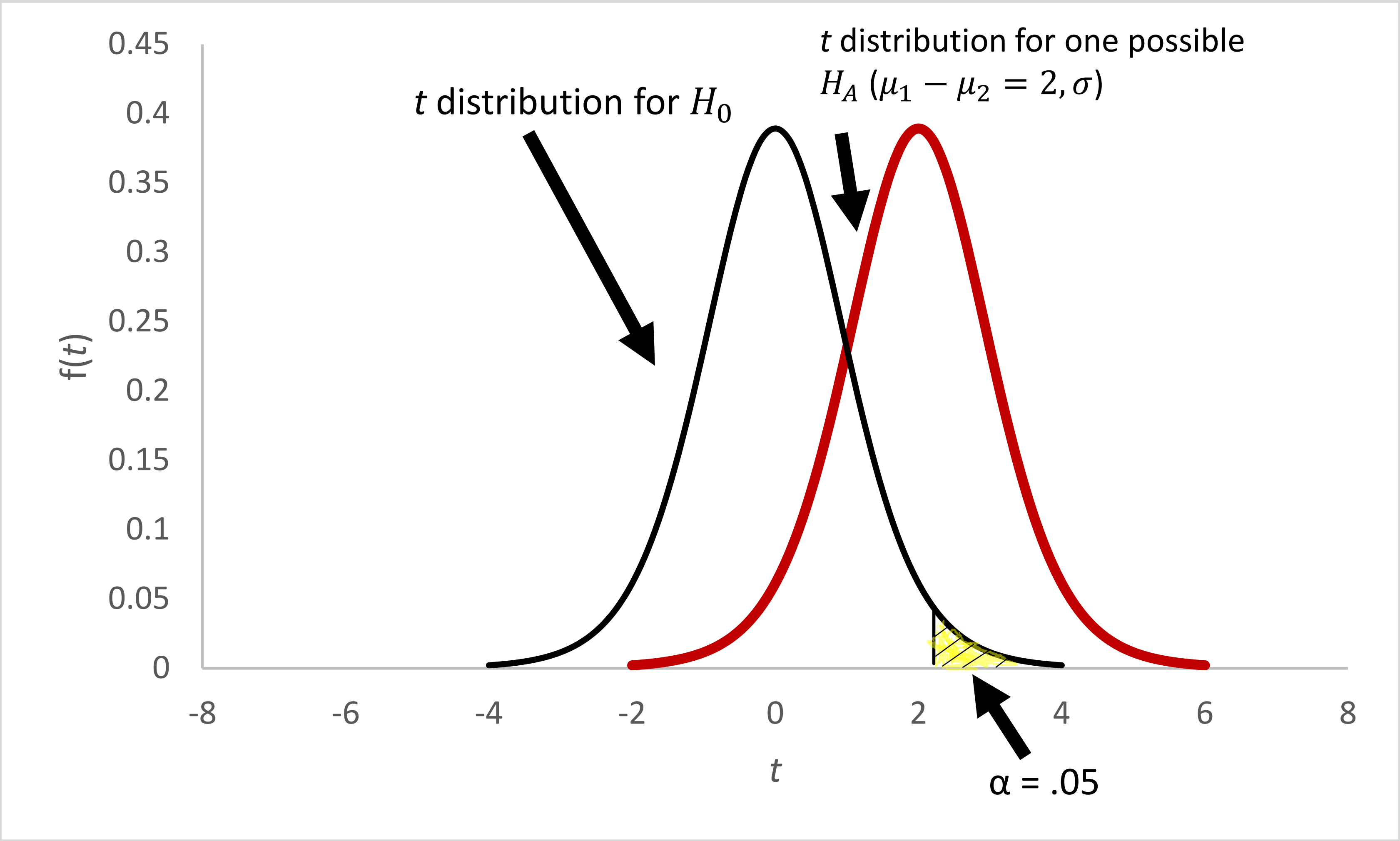
- A reference distribution used for hypothesis testing is the distribution of the test statistic when H_0 is true
  - This is called a central distribution
  - There is only one H_0, so only one central distribution

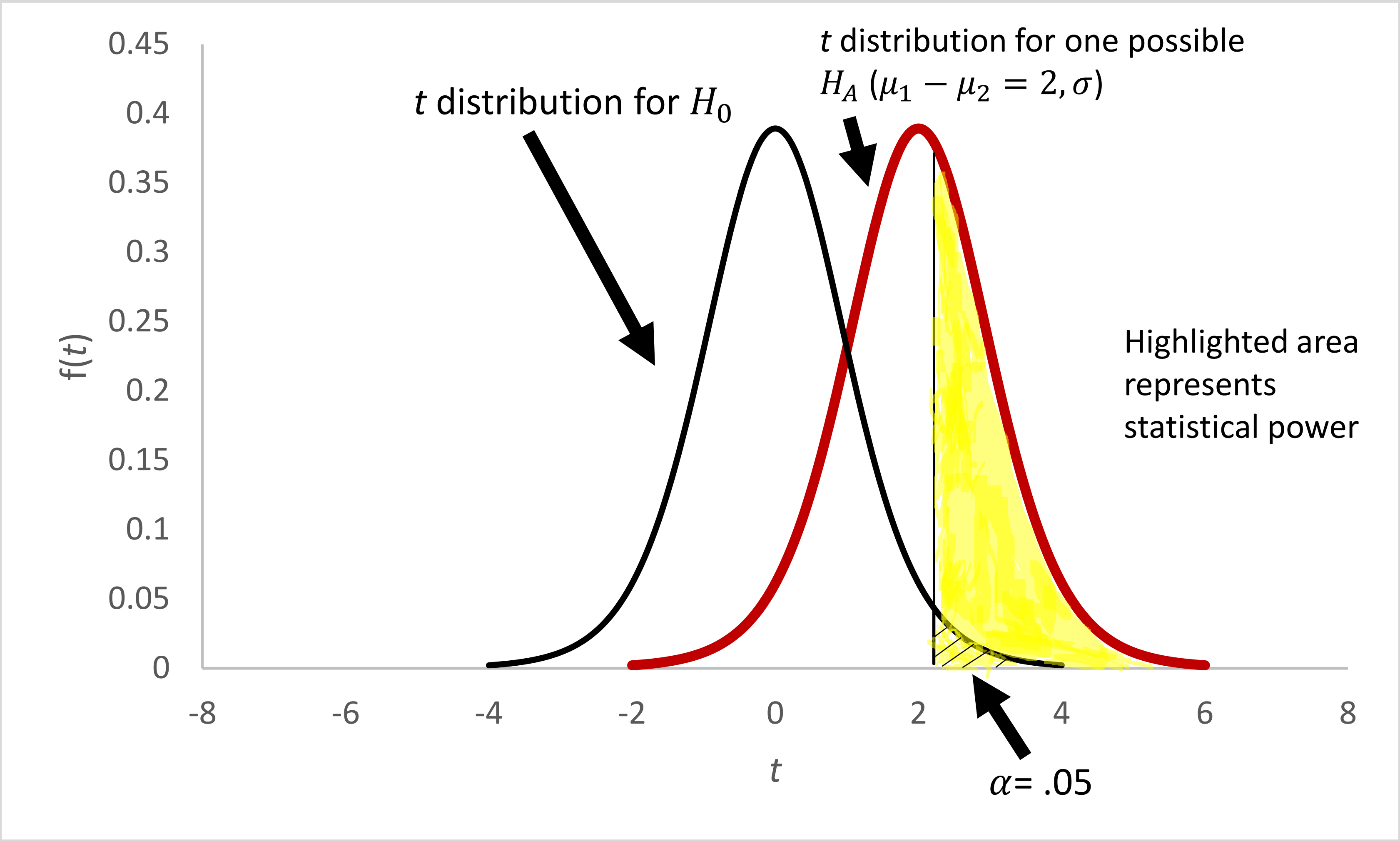
Statistical power 101

- The hypothesis that $\mu_1 - \mu_2 \neq 0$ is often called the alternative hypothesis, H_A or sometimes numbered H_1, H_2 etc.
- There are an infinite number of H_As
- WHY? Because $\mu_1 - \mu_2$ could equal any number

Statistical power 101

- Estimates of power are based on comparing the distribution of the test statistic when H_0 is true against the distribution for a particular H_A
  - A distribution of a test statistic for a particular H_A is called a noncentral distribution
- Power analysis requires specifying H_A, usually via an expression of anticipated effect size


t distribution for H_0
t distribution for one possible
H_A $(\mu_1 - \mu_2 = 2, \sigma)$
α = .05
f(t)
t
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-8
-6
-4
-2
0
2
4
6
8



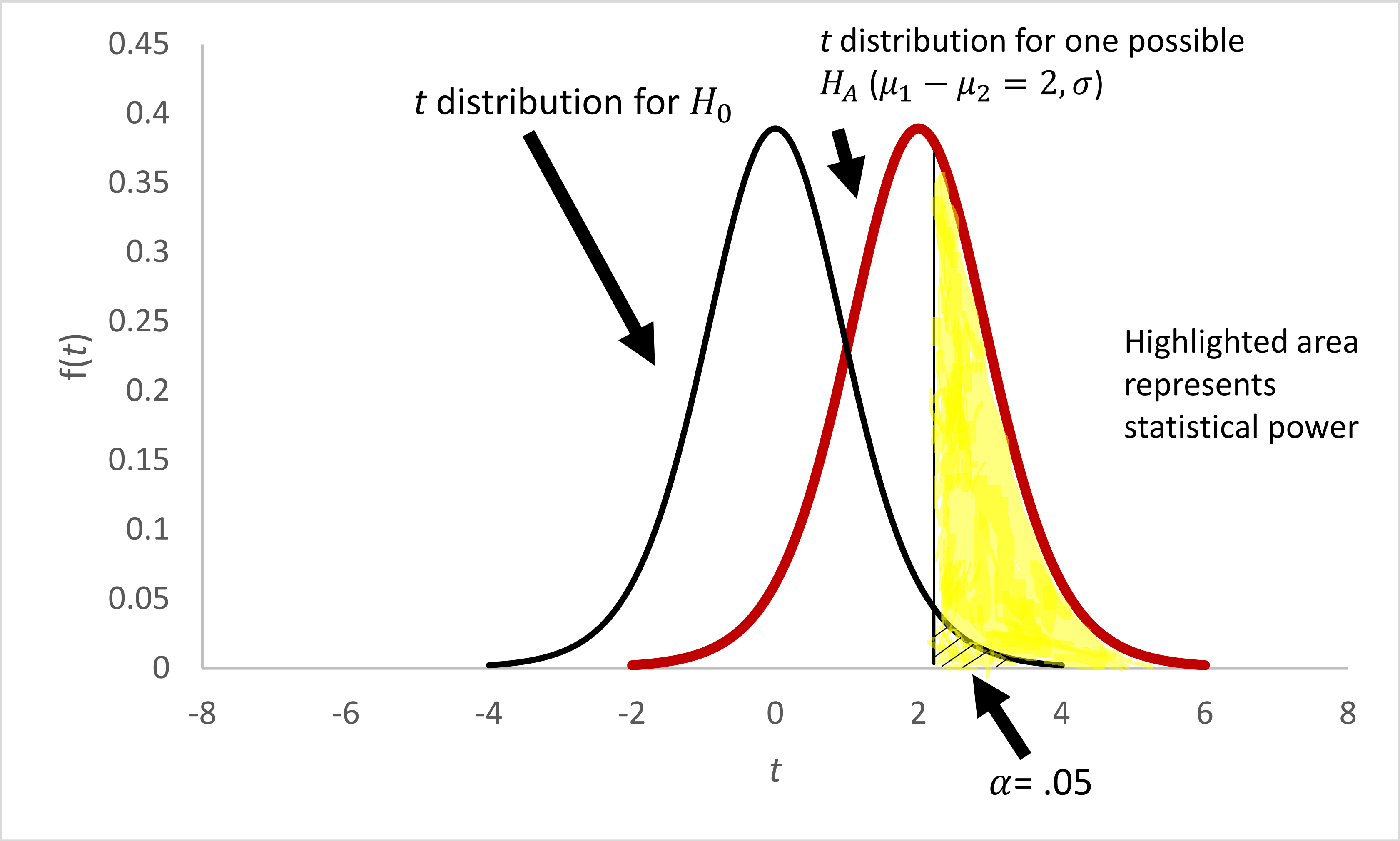
t distribution for H_0
t distribution for one possible
H_A $(\mu_1 - \mu_2 = 2, \sigma)$
Highlighted area
represents
statistical power
α= .05
f(t)
t
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-8
-6
-4
-2
0
2
4
6
8

Considerations affecting power

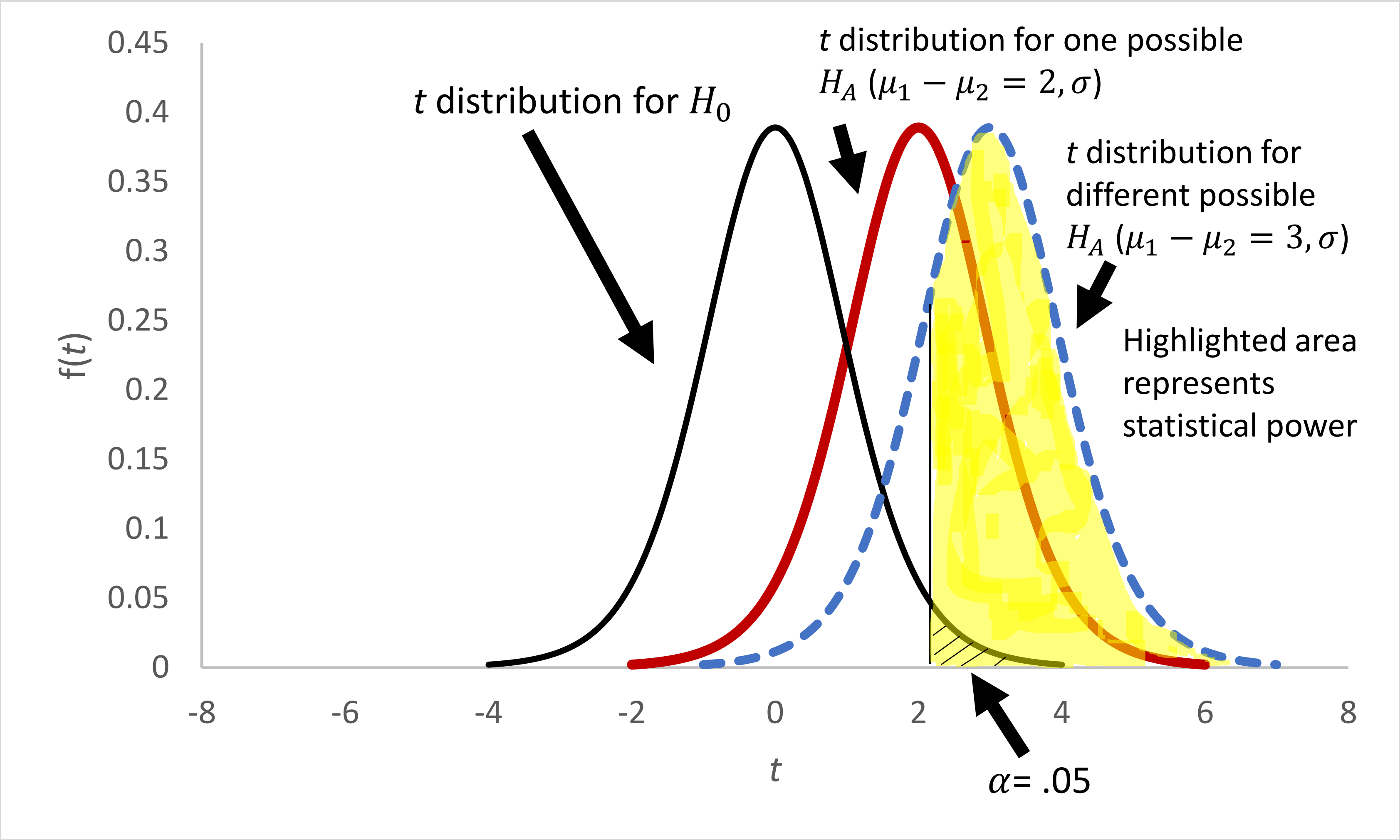
- Sample size (*N*)
  - All else being equal, larger *N* means more power
  - WHY? Larger sample size means smaller standard errors, more precise estimates, less “noise”

Considerations affecting power

- Effect size
  - All else being equal, larger effect size means more power
  - WHY? Larger effect size means less overlap between the test statistic distributions under H_0 and H_A



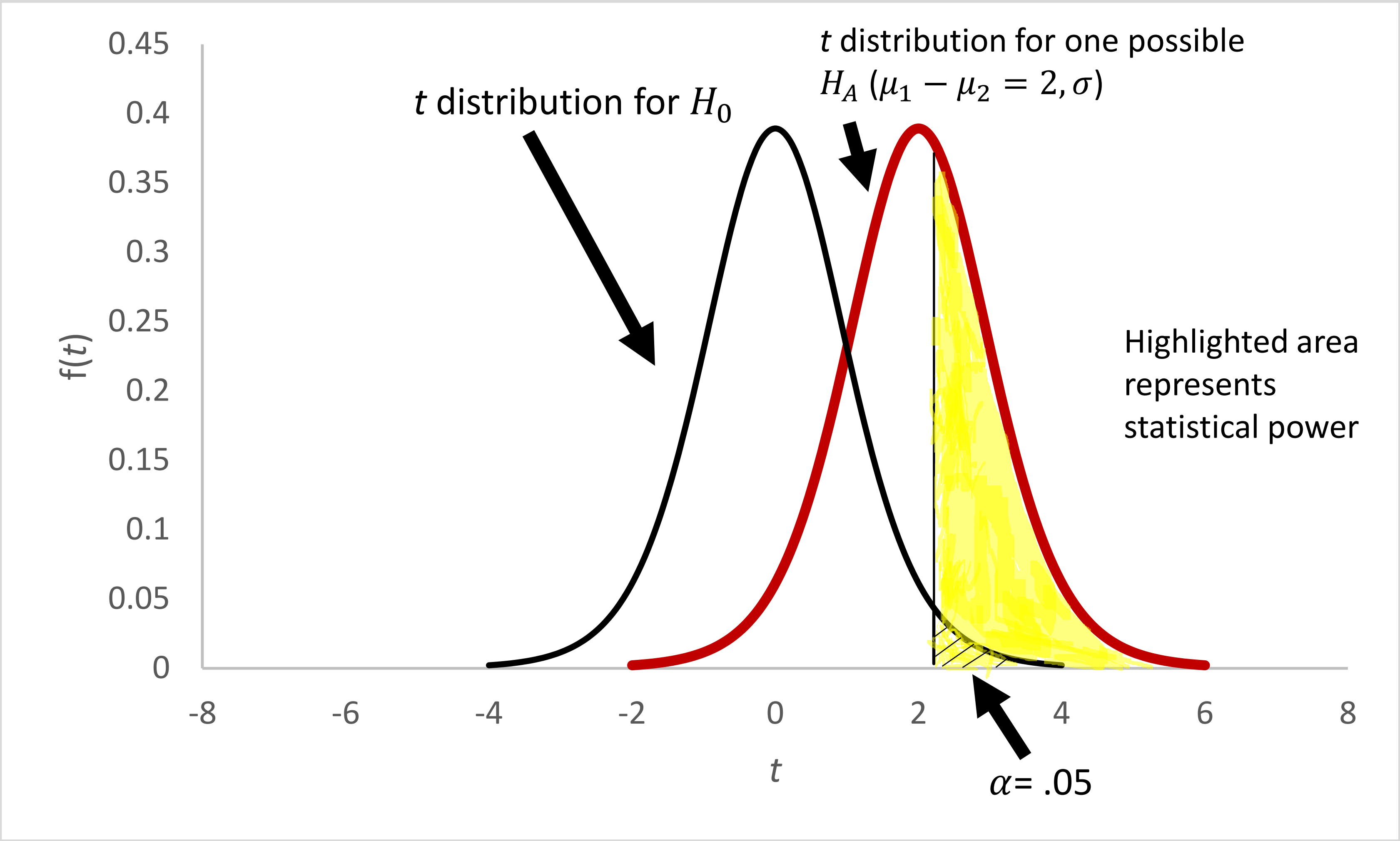
t distribution for H_0
t distribution for one possible H_A $(\mu_1 - \mu_2 = 2, \sigma)$
Highlighted area represents statistical power
α= .05
f(t)
t
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-8
-6
-4
-2
0
2
4
6
8



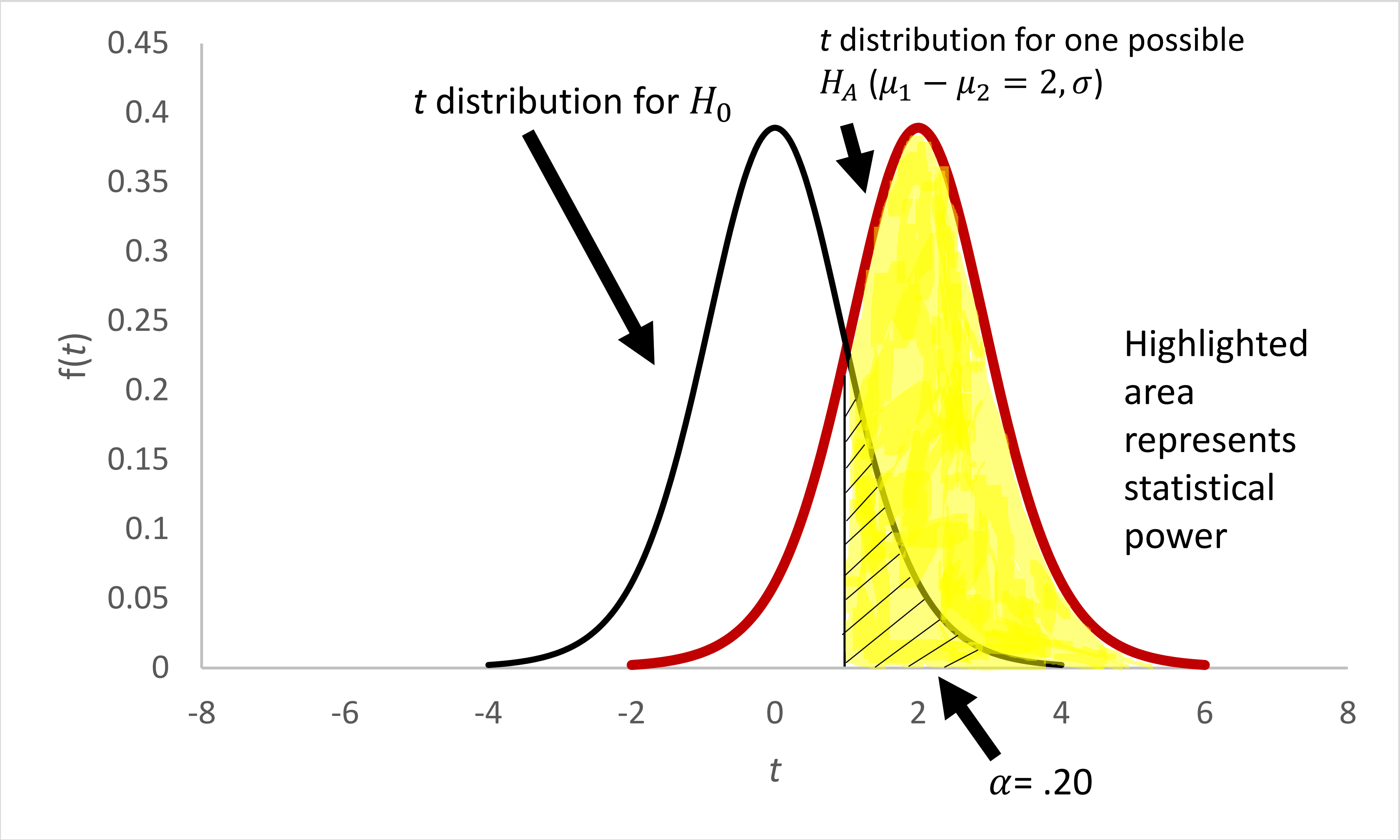
t distribution for H_0
t distribution for one possible H_A $(\mu_1 - \mu_2 = 2, \sigma)$
t distribution for different possible H_A $(\mu_1 - \mu_2 = 3, \sigma)$
Highlighted area represents statistical power
α= .05
f(t)
t
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-8
-6
-4
-2
0
2
4
6
8

Considerations affecting power

- Choice of α
  - All else being equal, larger alpha means more power
  - WHY? Easiest to see this graphically



t distribution for one possible H_A $(\mu_1 - \mu_2 = 2, \sigma)$
t distribution for H_0
Highlighted area represents statistical power
α= .05
f(t)
t
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-8
-6
-4
-2
0
2
4
6
8



t distribution for one possible H_A $(\mu_1 - \mu_2 = 2, \sigma)$
t distribution for H_0
Highlighted area represents statistical power
α= .20
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
f(t)
-8
-6
-4
-2
0
2
4
6
8
t

Considerations affecting power

- Sample size (*N*)
- Effect size
- Choice of α

- There is another consideration that comes up only when there is a hierarchical structure (nesting) in the data
- We will discuss this in a later lesson in this module

Expressing effect size for use in a power analysis

- One often-used effect size measure is Cohen's d (Cohen, 1988)

$$d = \frac{\bar{Y}_1 - \bar{Y}_2}{s}$$

- In factorial experiment terms, expresses effect size for a single two-level factor
- Essentially, standardized difference between two means

Expressing effect size

- Cohen established some benchmarks expressed in terms of d:
  - “small” = .2 to .3
  - “medium” = .5
  - “large” = .8 or greater

Expressing effect size of a regression coefficient

- Consider the regression of a variable Y on several X variables: X_1, X_2, etc. Denote the standardized regression coefficient associated with X_1 as β_1^*

- When the predictors are uncorrelated

$$\beta_1^* = \sqrt{\frac{r_{X_1Y}^2}{1 - r_{X_1Y}^2}}$$

Expressing effect size of a regression coefficient

$$\beta_1^* = \sqrt{\frac{r_{X_1Y}^2}{1 - r_{X_1Y}^2}}$$

- The Cohen benchmarks for β^* are
  - “small” = .1
  - “medium” = .25
  - “large” = .4 or greater

How to power factorial experiments

- You can choose α, level of power, and expected effect size, then establish necessary *N*
- OR
- Sometimes max *N* determined (e.g. by budget)
    - Minimum detectable effect

How to power a factorial optimization trial

- Remember: for a 2^k experiment, with equal *N*s, with effect coding, same regression weight size = same power irrespective of effect
- This means all else being equal, power is the same for regression coefficients corresponding to main effects and interactions

How to power a factorial optimization trial

- Decide on the effect size for each important effect. This can be either
  - The actual size you believe the effect will be OR
  - **The minimum effect that you care about**
    - The optimization trial is an “audition” for a role in your intervention
- You may have several different effect sizes corresponding to different factors. **Identify the smallest effect size**

How to power a factorial optimization trial

- Hint: Adding a covariate such as a pretest will often reduce sample size requirement
  - Do this only if correlation is substantial (above about $r = .5$)

In this lesson you learned how to:

- Understand the basics of powering a factorial experiment

In the next lesson you will learn how to:

- Distinguish between the conclusion-priority and decision-priority perspectives
- Discern whether the conclusion-priority or decision-priority perspective is appropriate in a given situation

References Cited

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.