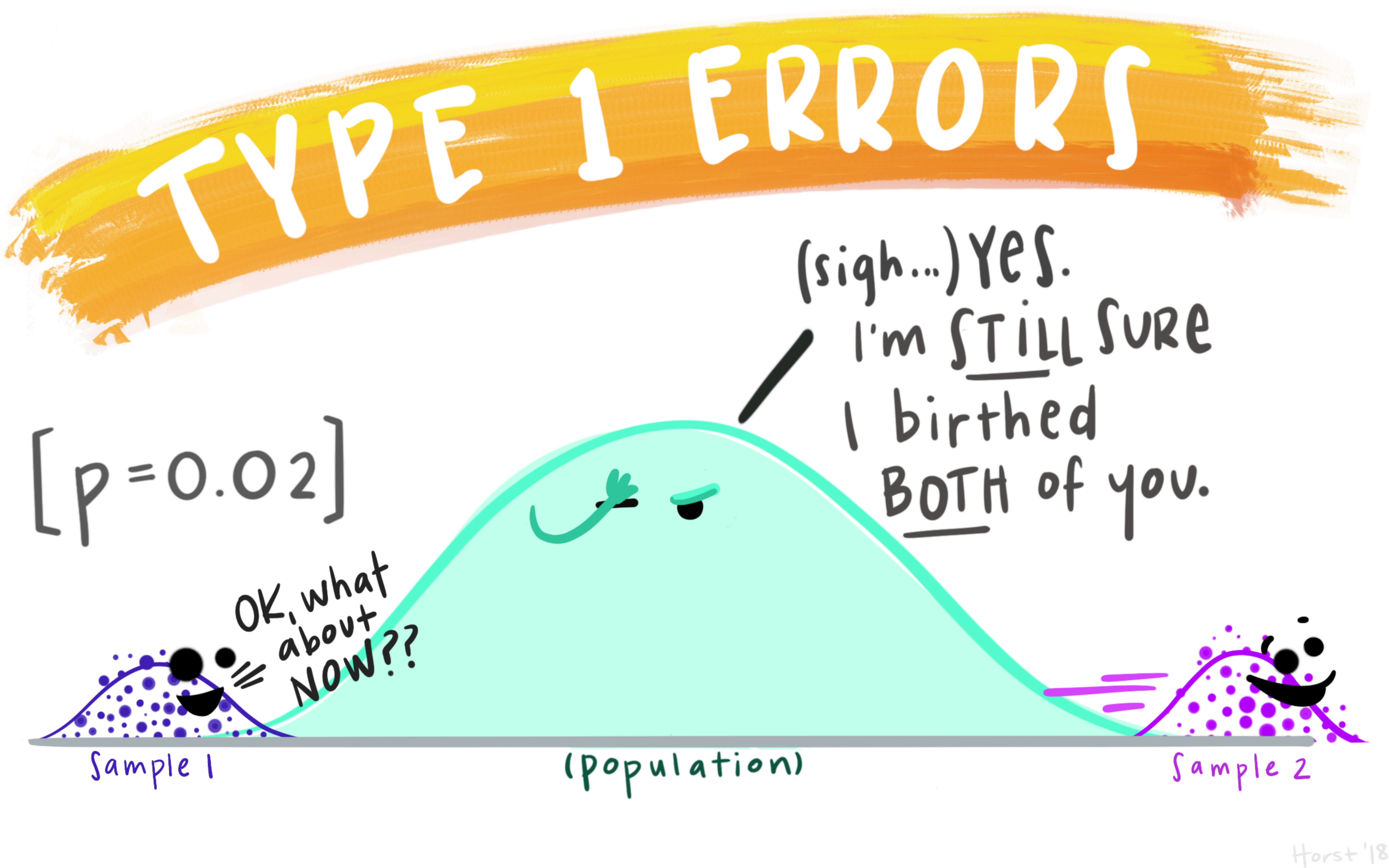
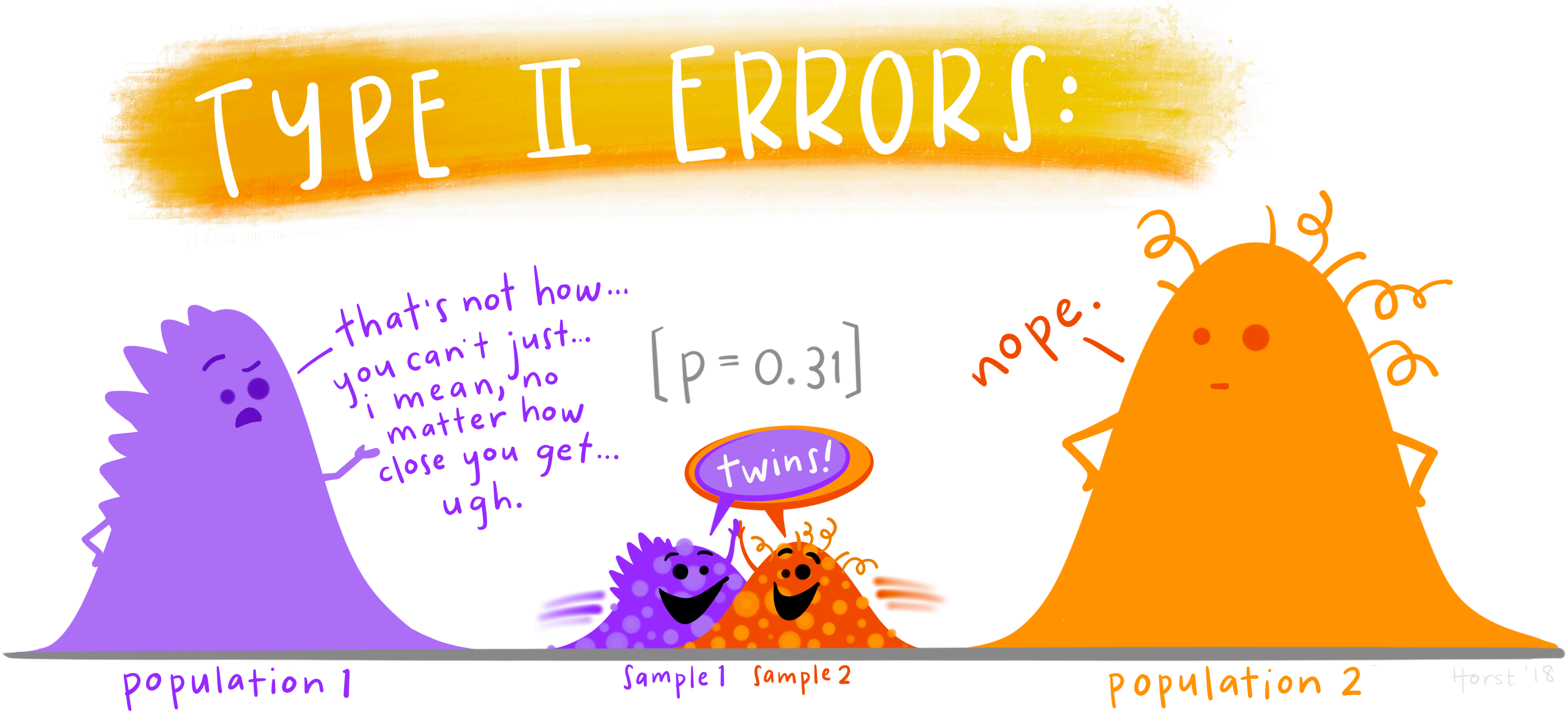
Errors and Power

Stat 250

Click here for PDF version


TYPE 1 ERRORS
[p=0.02]
(sigh...) Yes. I'm STILL sure I birthed BOTH of you.
OK, what about NOW??
Sample 1
(population)
Sample 2
Horst '18



TYPE II ERRORS:
that's not how... you can't just... i mean, no matter how close you get... ugh.
[p = 0.31]
twins!
nope.
population 1
sample 1
sample 2
population 2
Horst '18

Your turn 1

In a US court, the defendant is either innocent (H_0) or guilty (H_A)

a. What does a Type 1 Error represent in this context?

b. What does a Type 2 Error represent?

Your turn 2

A consumer protection agency is testing a sample of cans of tomato soup from a company. The company is presumed to be making a safe product, unless the consumer protection agency finds evidence that the average level of the chemical bisphenol A (BPA) in tomato soup from this company is greater than 100 ppb (parts per billion). If there is evidence that the average BPA concentration is unsafe, they will recall all the soup and sue the company.

Type I error rate

$$P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$$

AKA false positive rate

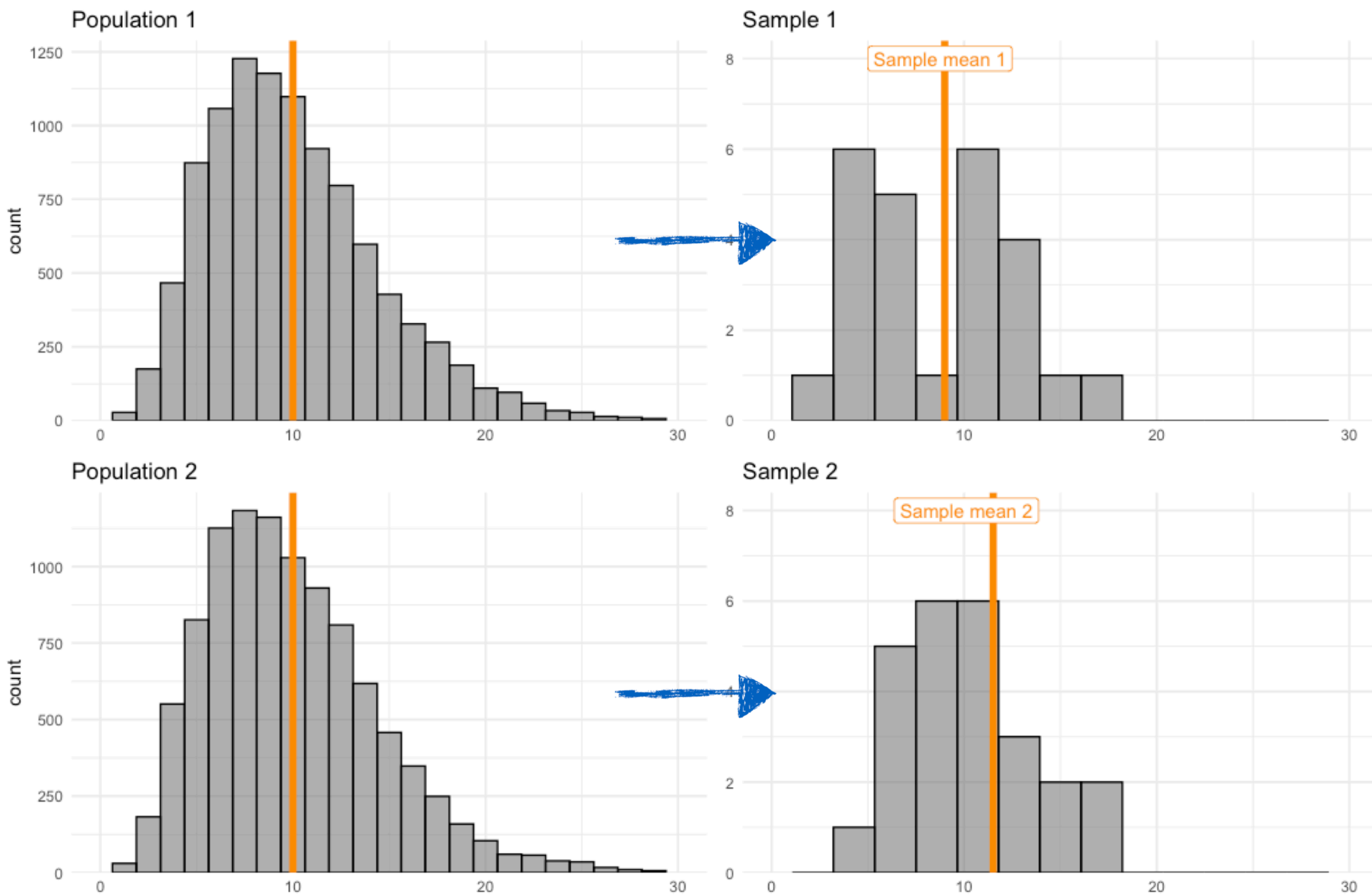
Where is this important?

- Decision rules/critical regions
- Sample size calculations

Your turn 3

- Let Y = # of successes out of $n = 20$ iid S/F trials; thus, $Y \sim \text{Binom}(n = 20, p)$.
- We wish to test the hypothesis $H_0 : p = .8$ versus the alternative, $H_a : p < .8$.
- Assume that the critical region $\{y \leq 12\}$ is used.
- Find α, the probability of making a type I error.

Simulation



Population 1
1250
1000
750
500
250
0
count
0
10
20
30
Sample 1
8
6
2
0
Sample mean 1
0
10
20
30
Population 2
1000
750
500
250
0
count
0
10
20
30
Sample 2
8
6
2
0
Sample mean 2
0
10
20
30

Two-sample hypothesis test

$H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} = -1.83$$

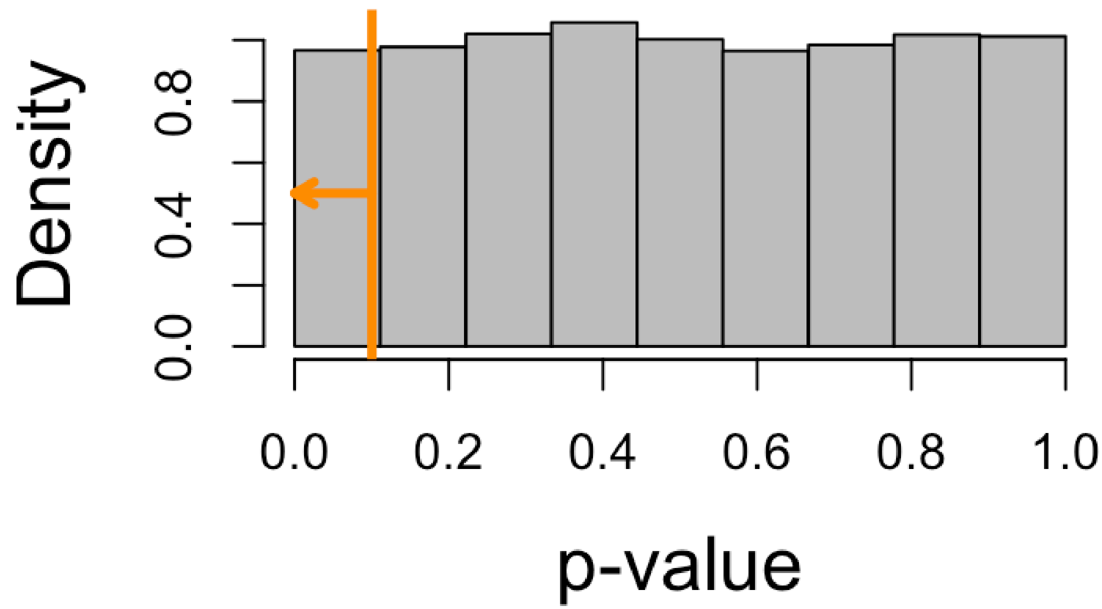
```
t <- t.test(SAMPLE1, SAMPLE2)
t$p.value
```

```
[1] 0.02594196
```

So if we reject when $p\text{–value} < 0.1$, we would make a **Type I error**

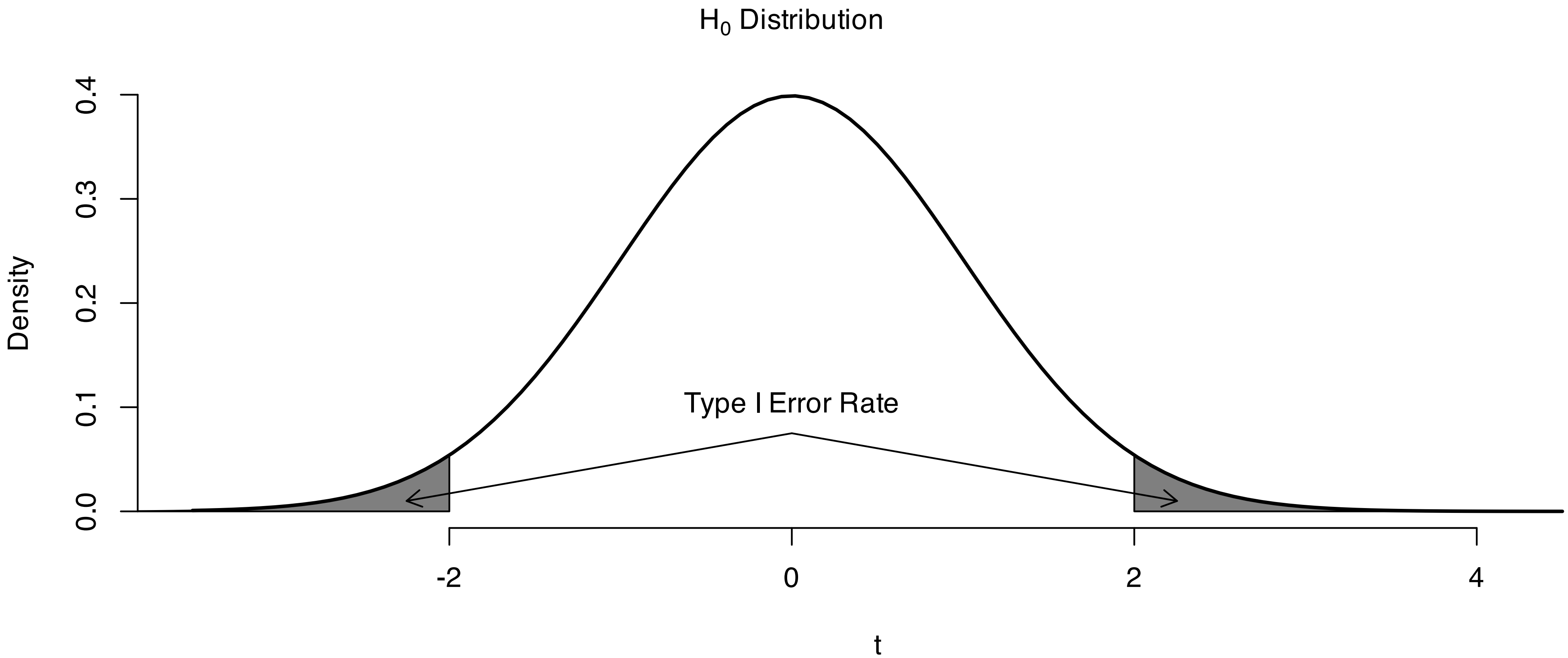
Distribution of p-values

p-values are uniformly distributed under H_0



True $\mathrm{H}_0 \implies$ 10% of the p-values are less than 0.1

Type I error



H0 Distribution
Density
0.4
0.3
0.2
0.1
0.0
Type I Error Rate
-2
0
2
4
t

Type II error rate

$$P(\text{Type II error}) = P(\text{Do not reject } H_0 | H_0 \text{ is false}) = \beta$$

AKA false negative rate

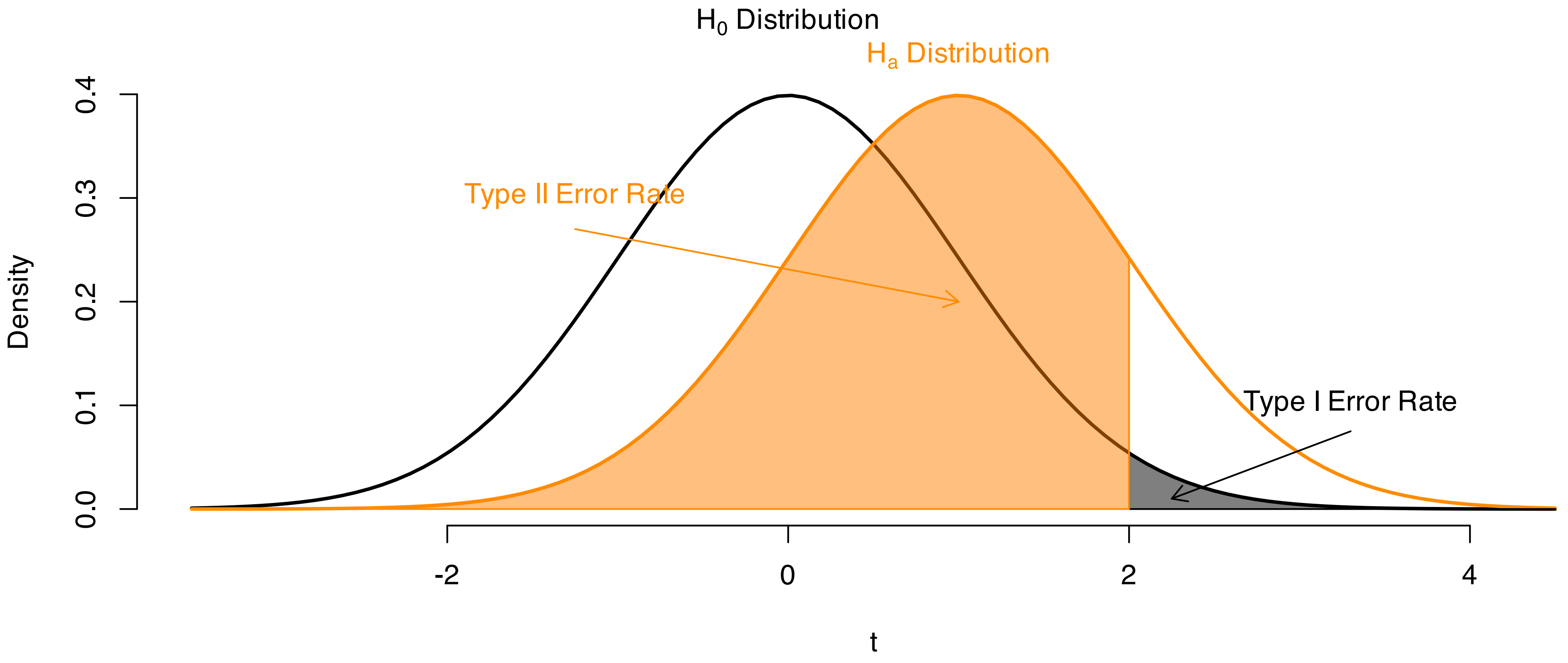
Where is this important?

- Power
- Sample size calculations

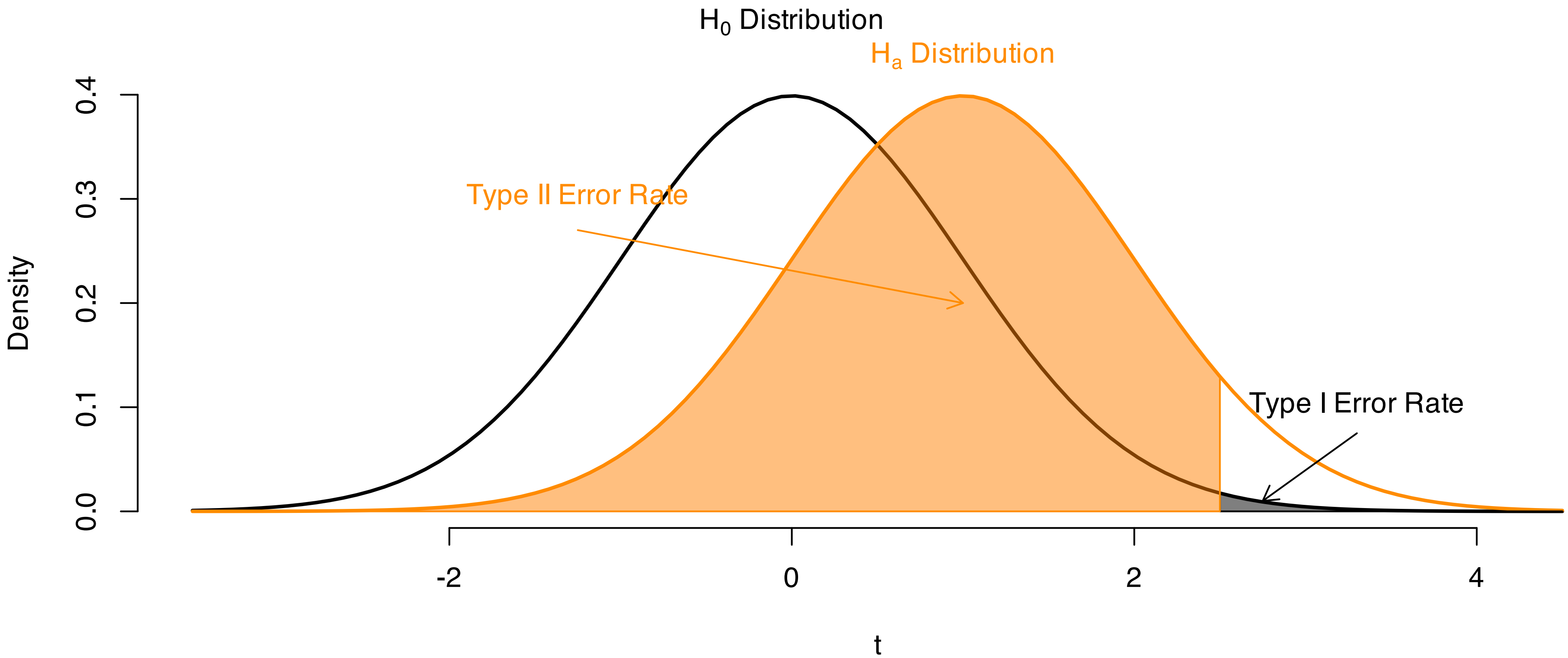
Your turn 4

- Let Y = # of successes out of $n = 20$ iid S/F trials; thus, $Y \sim \text{Binom}(n = 20, p)$.
- We wish to test the hypothesis $H_0 : p = .8$ versus the alternative, $H_a : p < .8$.
- Assume that the critical region $\{y \leq 12\}$ is used and that the truth is that $p = 0.6$ (so H_0 is false).
- Calculate β.

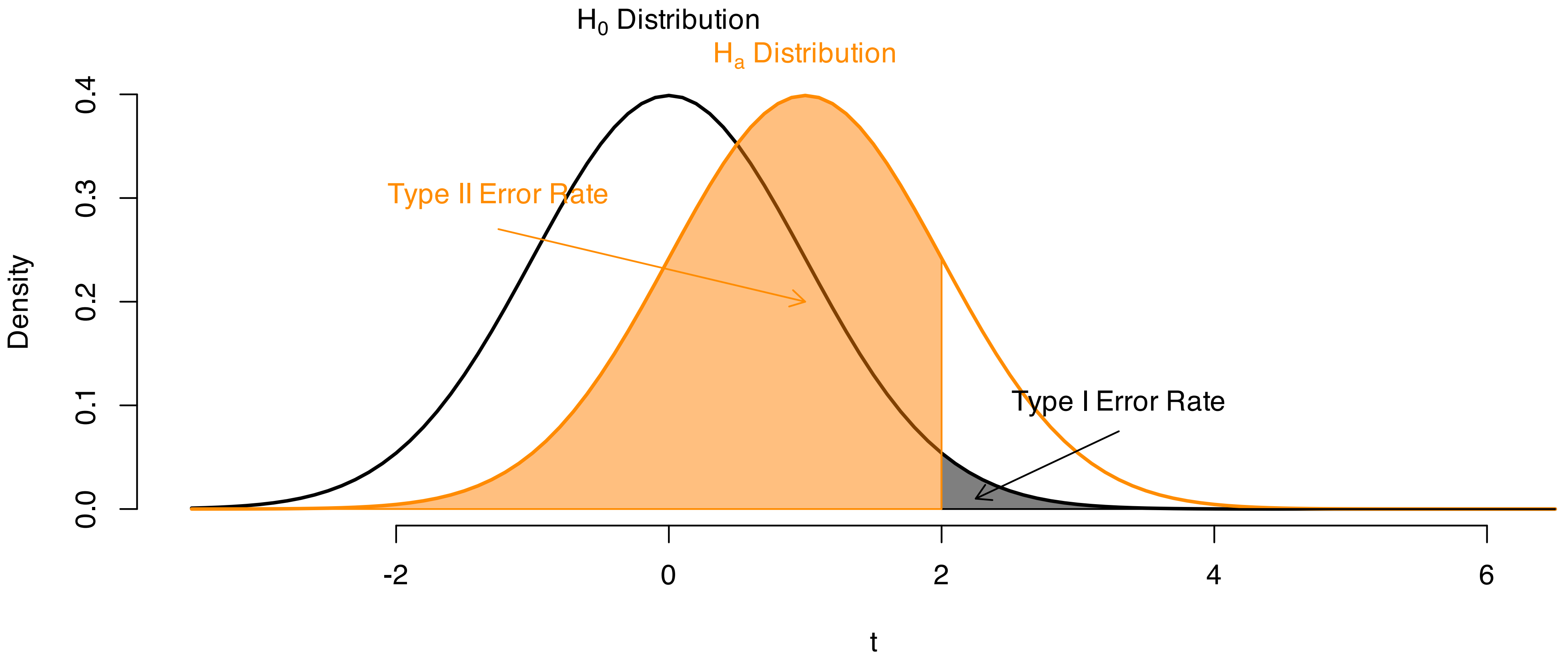
Type I and type II errors



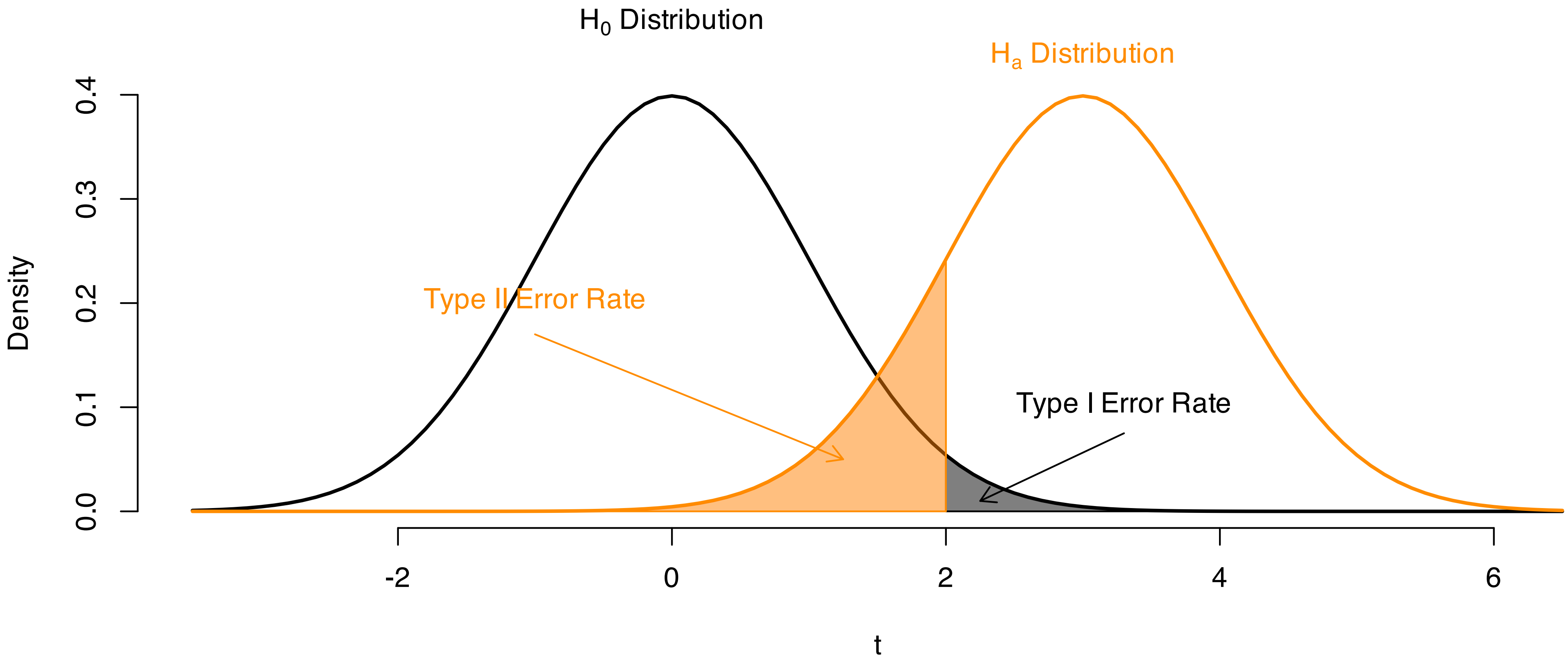
Type I and type II errors



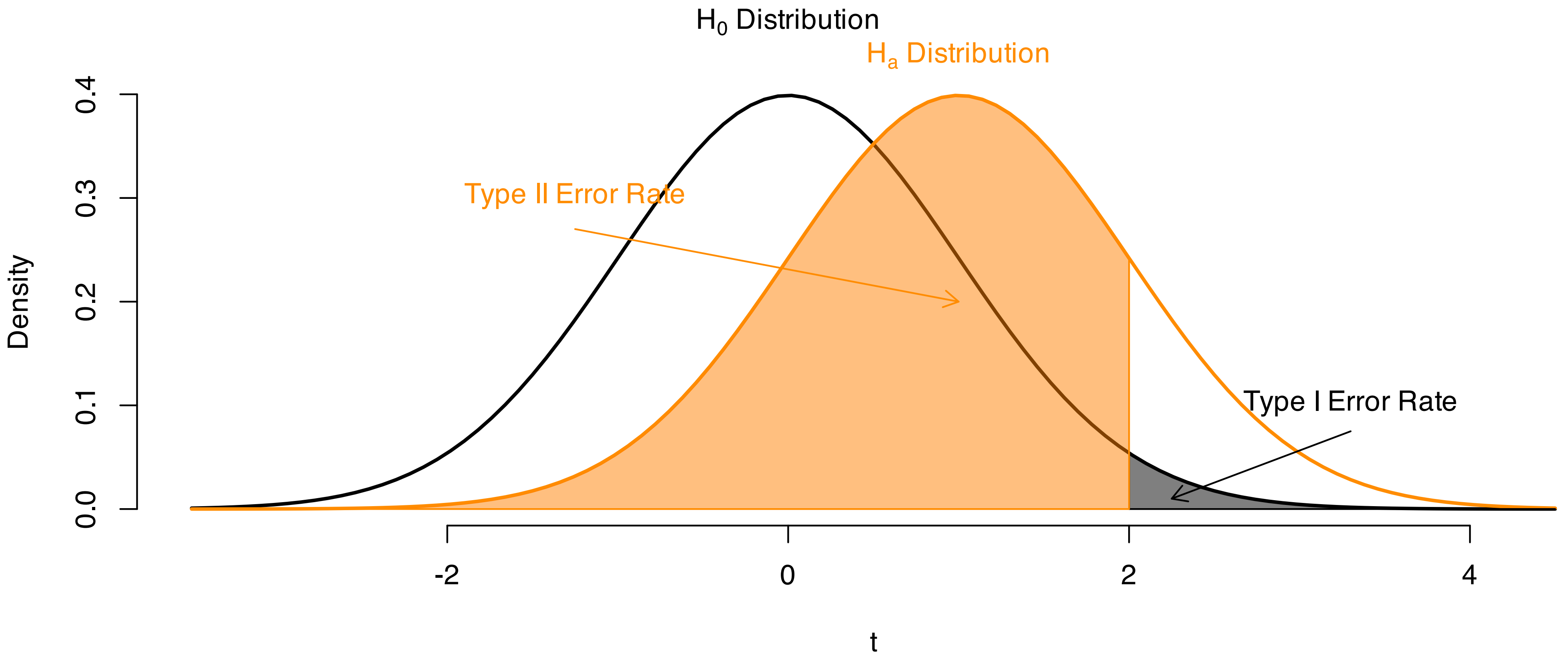
Effect size and type II errors



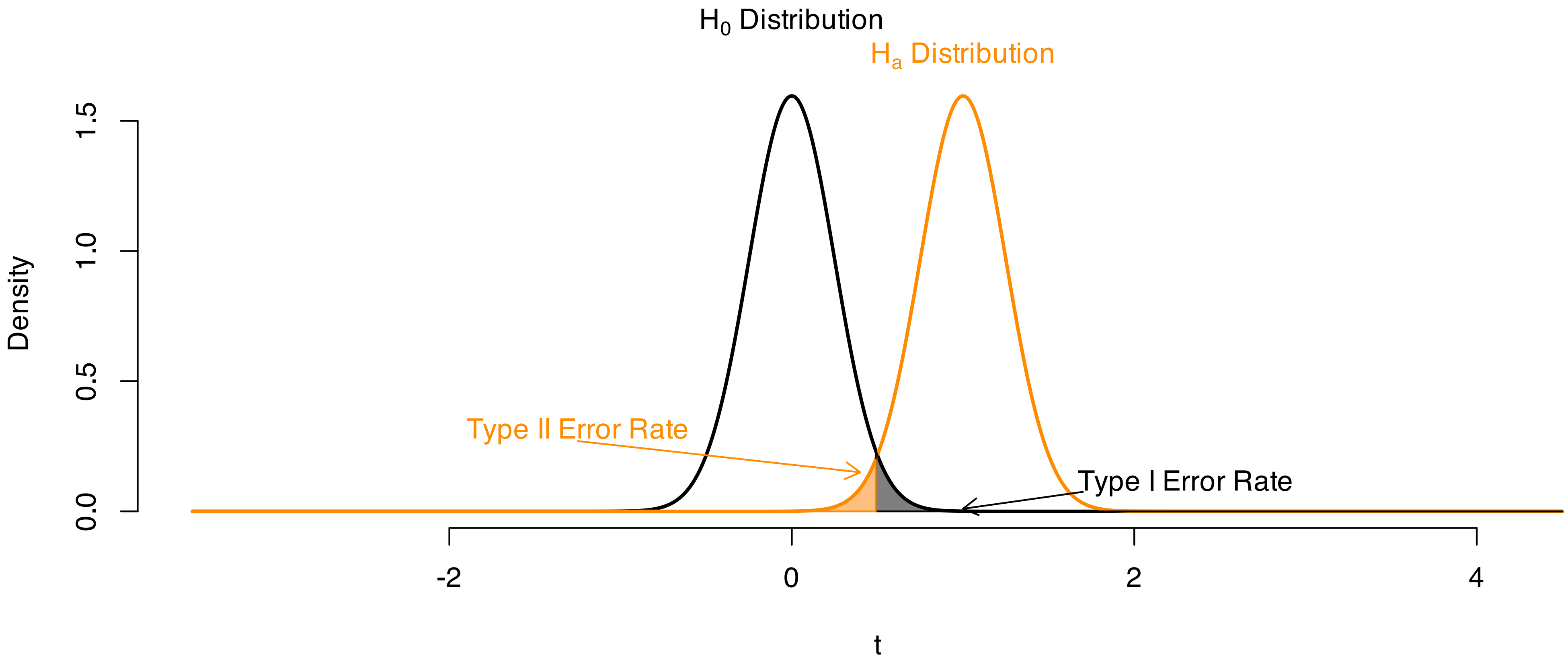
Effect size and type II errors



SE and type II error rate



SE and type II error rate



Power

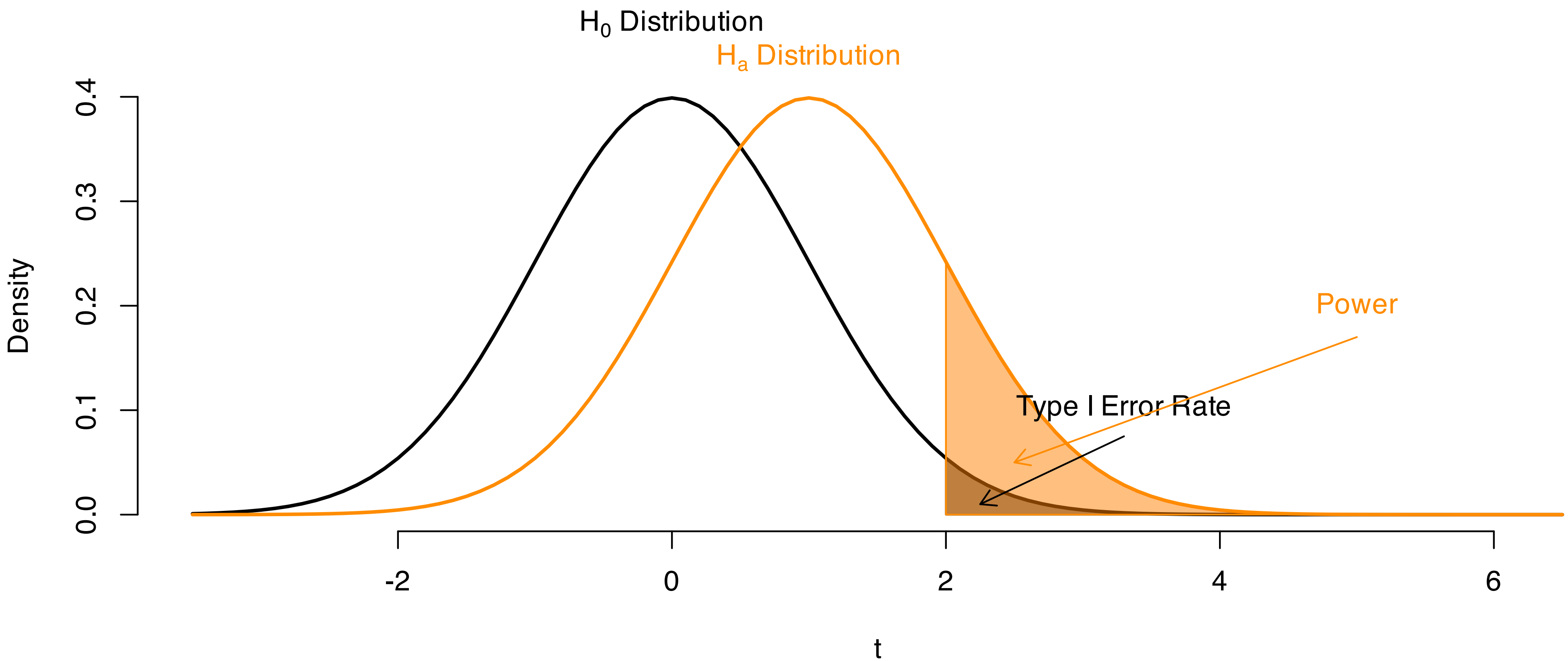
$$\text{Power} = P(\text{Reject } H_0 | H_0 \text{ is false}) = 1 - \beta$$

- Statisticians often talk about the power of a test rather than type II error
- Good tests have high power

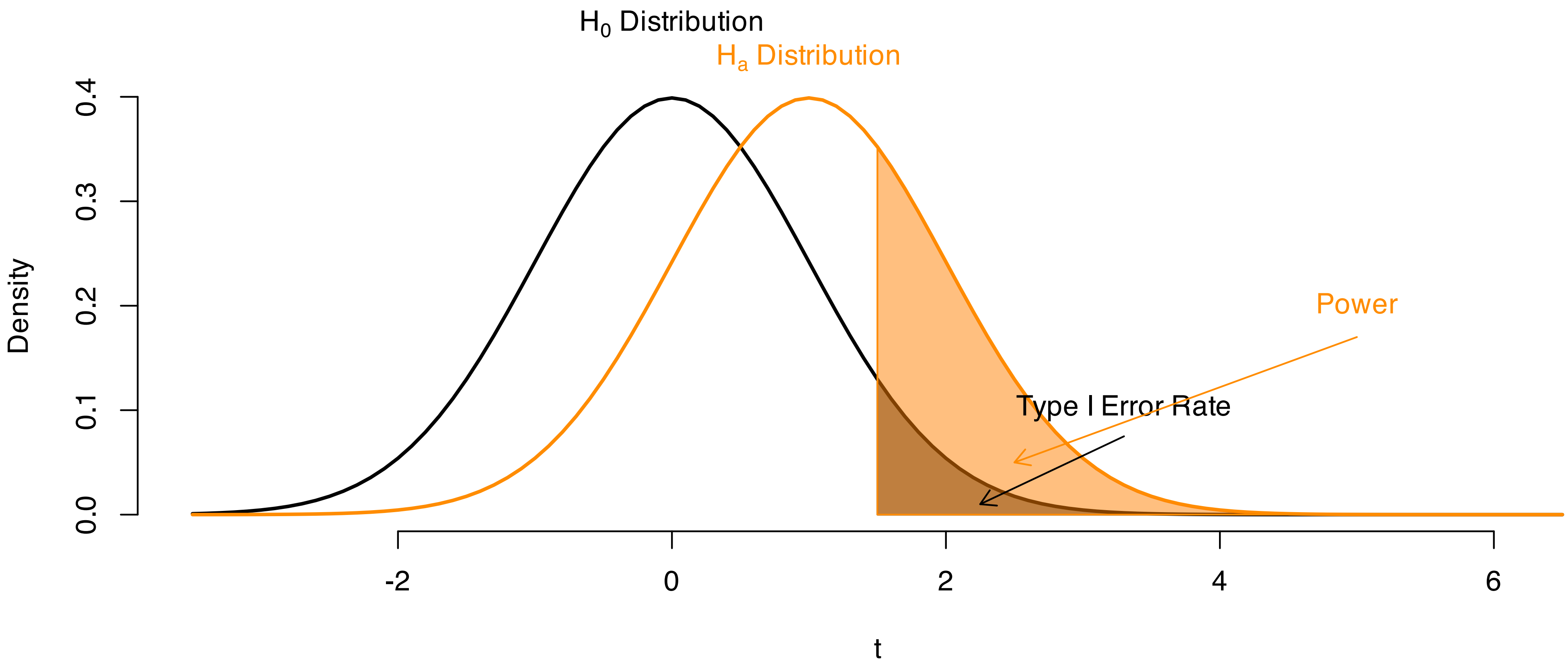
Your turn 5

- Let Y = # of successes out of $n = 20$ iid S/F trials; thus, $Y \sim \text{Binom}(n = 20, p)$.
- We wish to test the hypothesis $H_0 : p = .8$ versus the alternative, $H_a : p < .8$.
- Assume that the critical region $\{y \leq 12\}$ is used and that the truth is that $p = 0.6$ (so H_0 is false).
- Calculate the power of the test.

Power and α

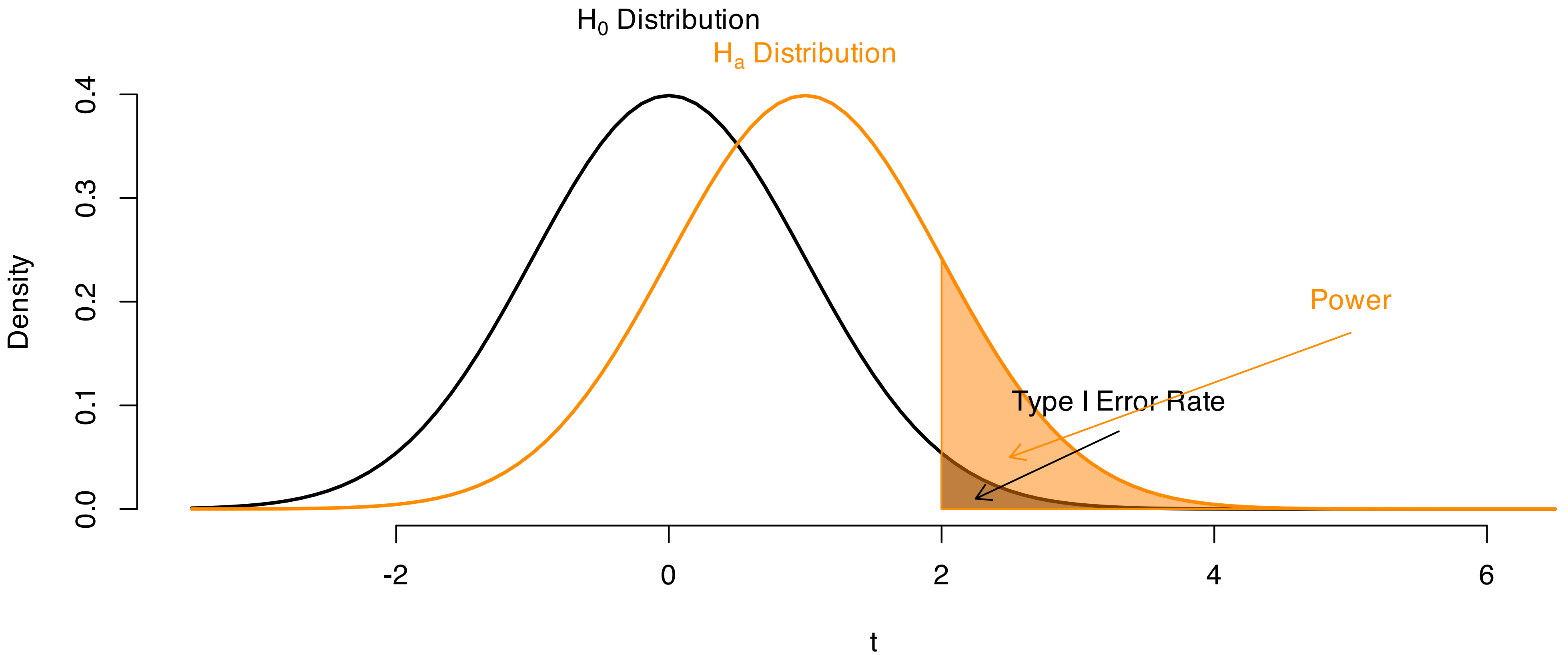


Power and α

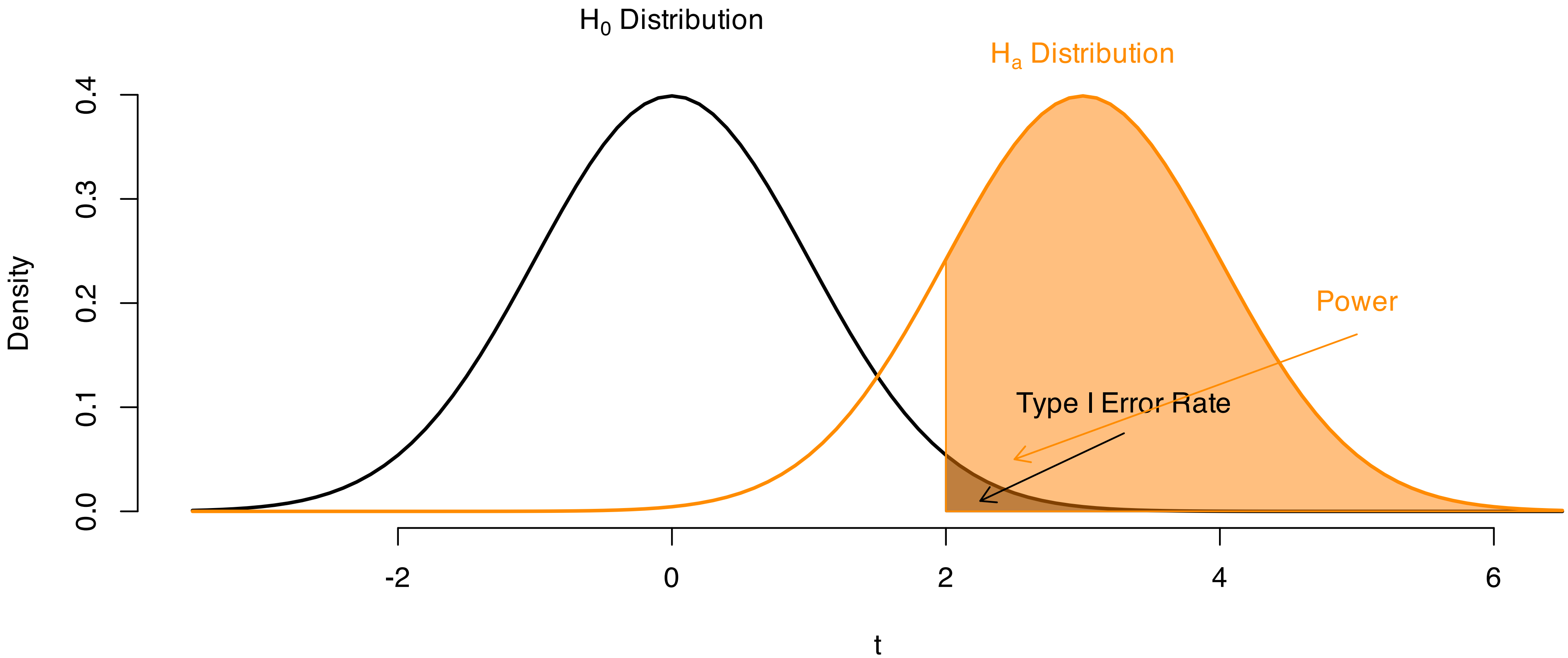


H_0 Distribution
H_a Distribution
Density
0.0
0.1
0.2
0.3
0.4
Power
Type I Error Rate
-2
0
2
4
6
t

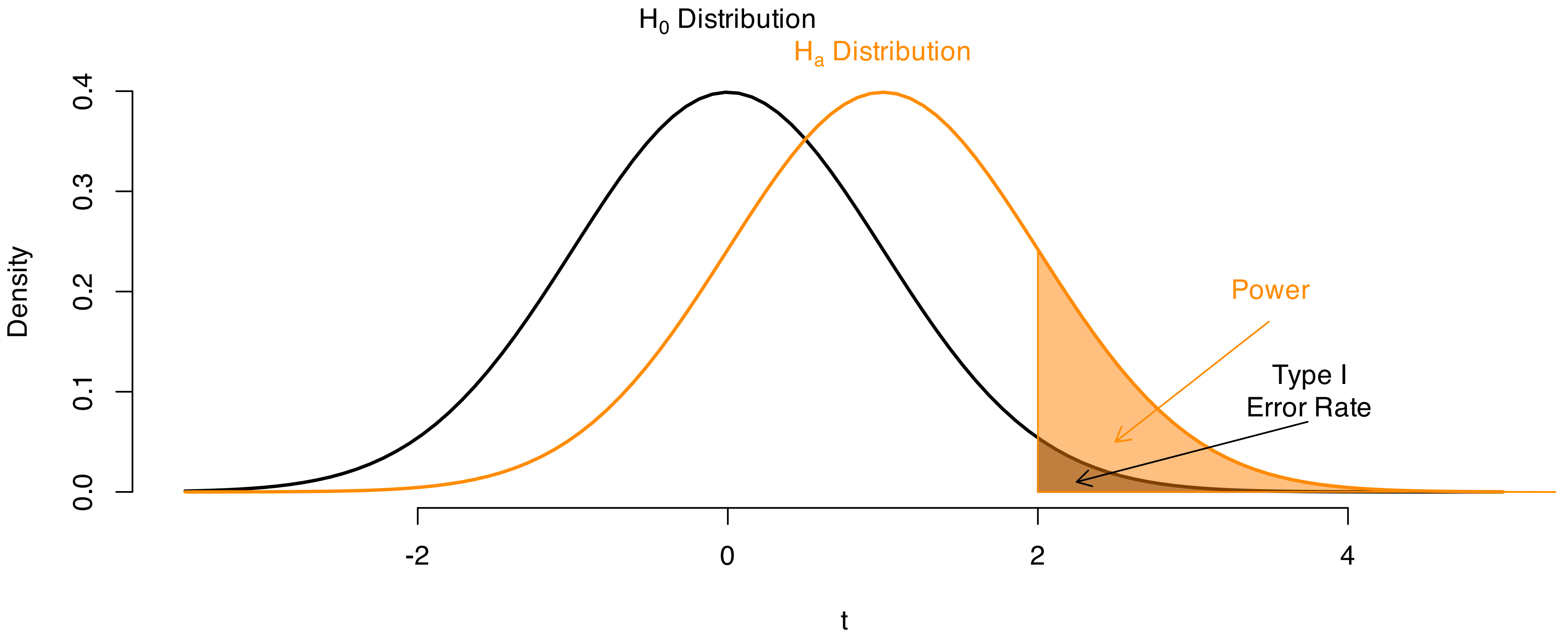
Power and effect size



Power and effect size

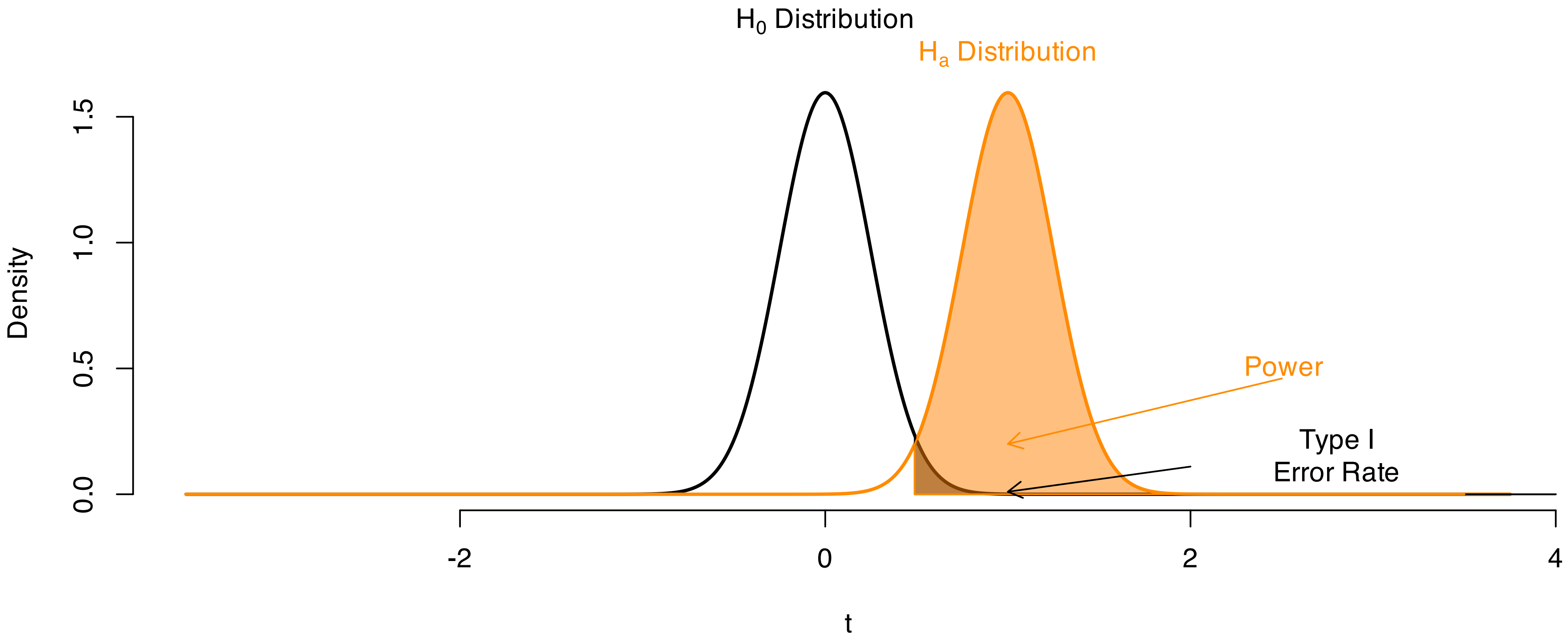


Power and SE



H_0 Distribution
H_a Distribution
Power
Type I
Error Rate
Density
0.0
0.1
0.2
0.3
0.4
-2
0
2
4
t

Power and SE



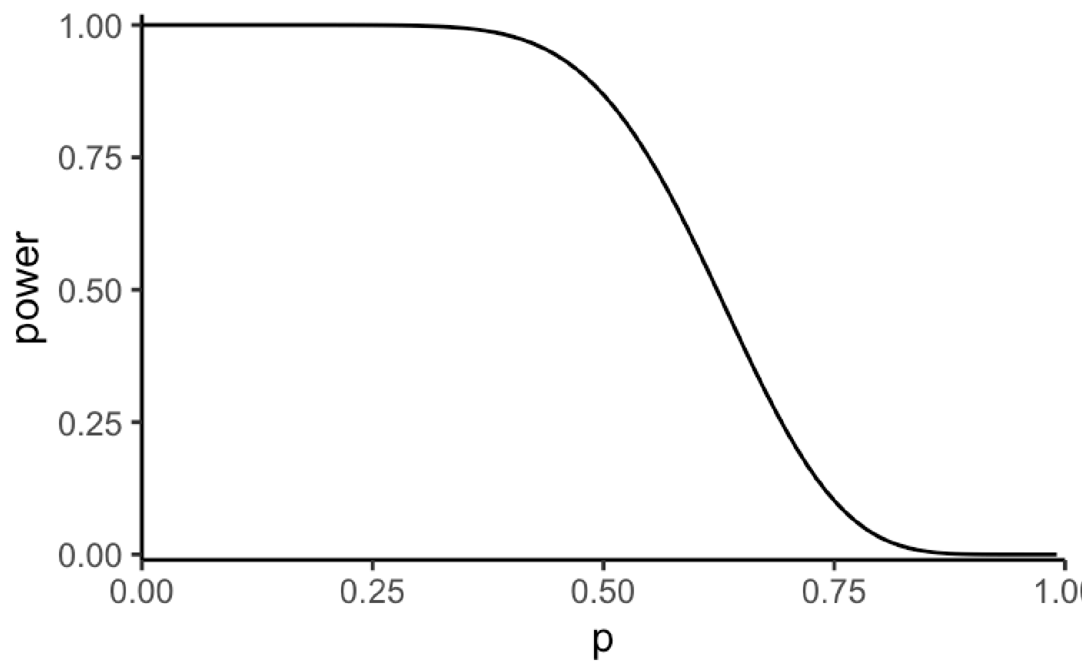
H_0 Distribution
H_a Distribution
Density
1.5
1.0
0.5
0.0
Power
Type I
Error Rate
-2
0
2
4
t

Power curves

Plot of power against different alternatives for the parameter

Example:

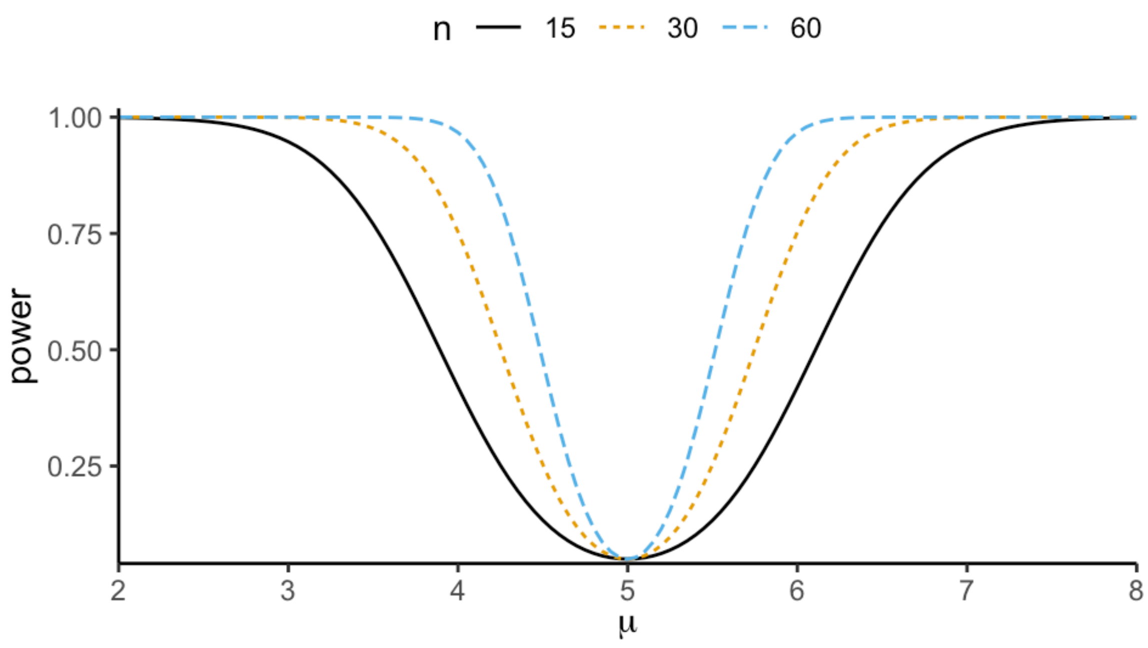
- $H_0 : p = .8$ vs. $H_A : p < .8$
- $Y \sim \mathrm{Binom}(n = 20, p)$
- critical region $\{y \leq 12\}$



Power curves

One-sample t-test

- $H_0 : \mu = 5$ vs. $H_a : \mu \neq 5$
- Assuming $s = 2$



Test planning

If planning a hypothesis test **before** data collection, you can determine the necessary sample size to achieve a specified power when using a set significance level/rejection region.

Example: Let $Y_i \overset{iid}{\sim} N(\mu, 5^2)$. We wish to test $H_0 : \mu = 7$ vs. $H_A : \mu > 7$ at the $\alpha = 0.05$ level. What is the smallest sample size such that the test has power at least .80 when $\mu = 8$?