Lecture 23
Tests for μ and p,
Types of errors in hypothesis tests, Power

Warm Up

- Using fracking, the U.S has become the largest oil produce in the world. Despite its economic benefits, fracking has become controversial due to its environmental impacts. Survey was conducted to quantity public opinion about fracking. The survey interviewed 1,353 Americans and found that 637 reported being against fracking. The researchers are interested in whether or not there is evidence that most Americans are opposed to fracking.

Conduct a significance test at the $\alpha = 0.05$ significance level to determine if there is evidence for a majority opinion against fracking. Use the five steps for a hypothesis test

Warm Up

- A government agency is interested in understanding the average annual income of households in a particular region to inform economic policies. The agency hypothesizes that the mean annual income of households in this region is \$50,000. To test this, they collect a random sample of 100 households from the region and estimate the annual income to be 45,323 with a standard deviation of \$13,121. Conduct a hypothesis test at the α = 0.05 significance level to determine if the annual income is significantly different than the agencies hypothesis for this region.

Significance tests are less useful than confidence intervals

- Significance tests have been overemphasized in practice
- A significance test only tells you whether or not a given parameter value in the null hypothesis (such as $\mu_0 = 0$, or $p_0 = 0.5$) is plausible given the data.
- When a P-value is small, it indicates the value specified by the null is not plausible but tells us little else about the possible values of the parameter.
- A confidence interval is more informative because it tells us the entire set of plausible values

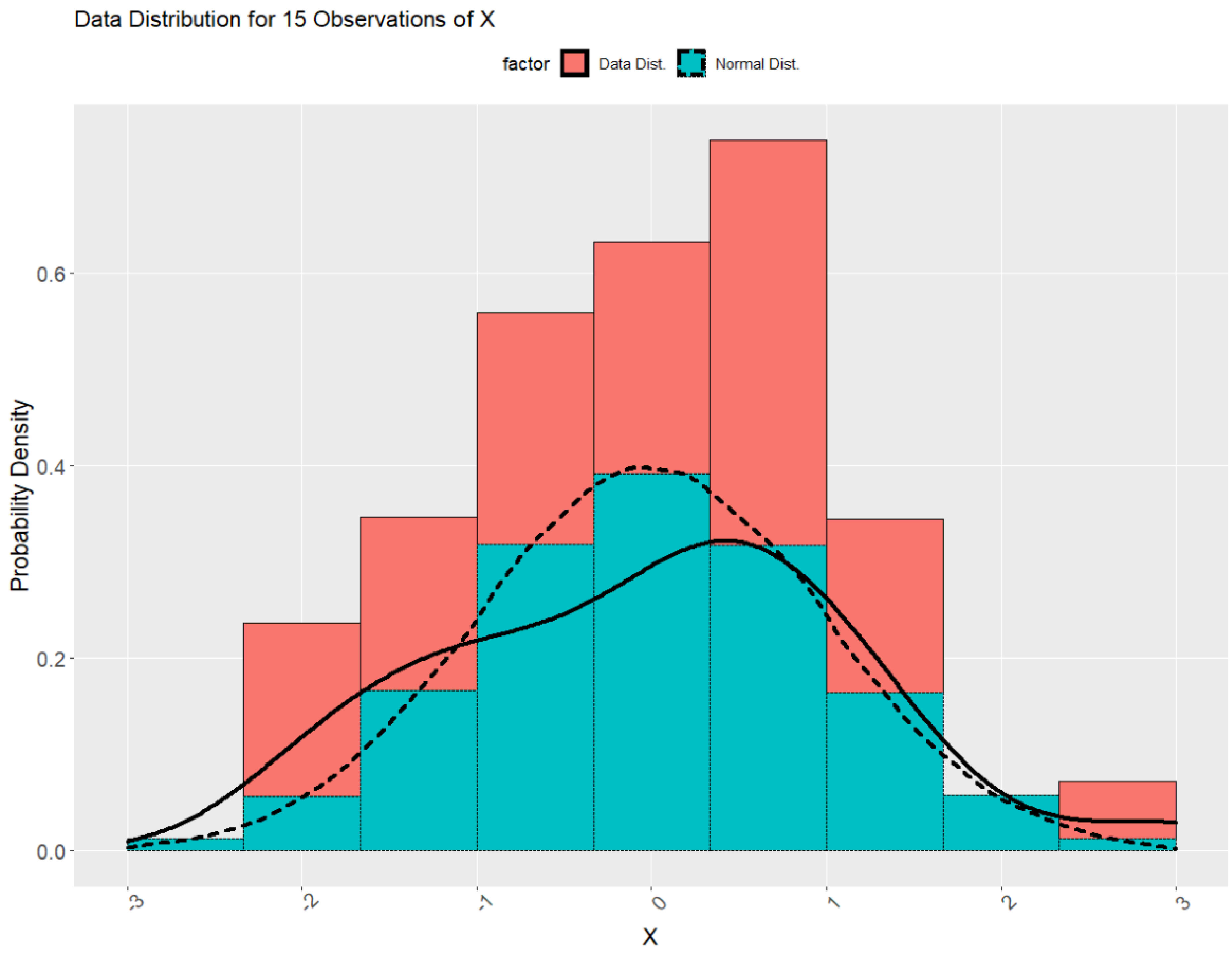
Checking Assumptions

- For a test concerning p a simple check to ensure

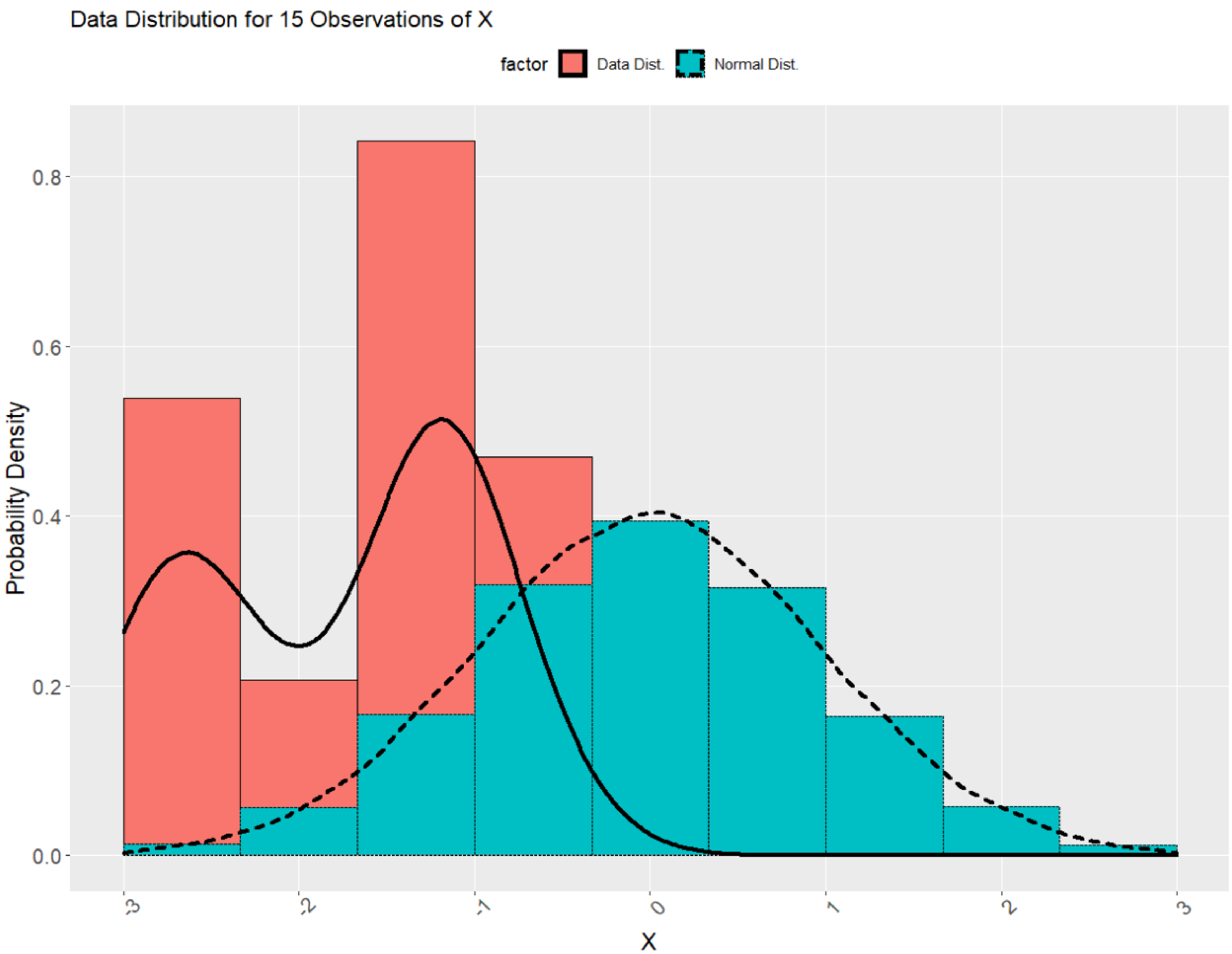
$np \geq 15$ and $n(1-p) \geq 15$ is sufficient to meet the assumptions

- For μ a histogram of the data distribution of x is an easy way to determine if the population distribution of x is approximately normal
  - x represents the variable from the data that we are conducting a hypothesis test on

- The normality assumption about the population distribution of x is most important when n is small and the test is one sided
  - Two-sided tests are more robust to deviations from normality



Data Distribution for 15 Observations of X
factor
Data Dist.
Normal Dist.
Probability Density
0.0
0.2
0.4
0.6
-3
-2
-1
0
1
2
3
X



Data Distribution for 15 Observations of X
factor
Data Dist.
Normal Dist.
Probability Density
0.0
0.2
0.4
0.6
0.8
-3
-2
-1
0
1
2
3
X

Example 1: Test for a population mean

Example: Consider the following data from a study of the volume of the left hippocampus for twin pairs discordant for schizophrenia.

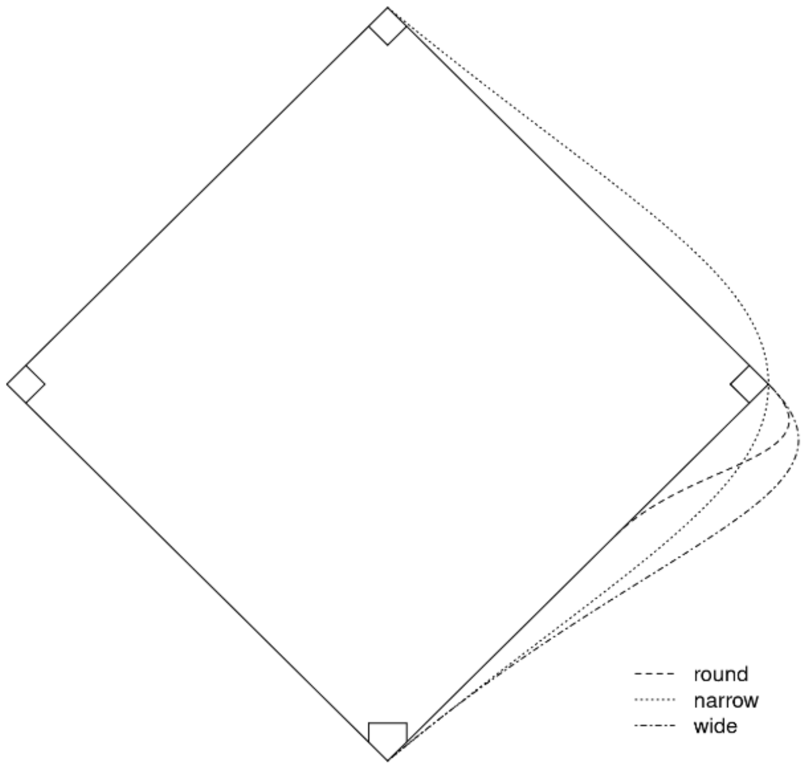
	Twin		
Pair	**Unaffected**	**Affected**	**Difference**
1	1.94	1.27	0.67
2	1.44	1.63	-0.19
3	1.56	1.47	0.09
4	1.58	1.39	0.19
5	2.06	1.93	0.13
⋮	⋮	⋮	⋮
15	2.08	1.97	0.11

The mean difference from the sample is $\bar{x} = 0.2$ cubic centimeters, and the standard deviation from the sample is $s = 0.24$ cubic centimeters. Is this result statistically significant at a signficiance level of $\alpha = 0.05$?

Another Example

- Each of 22 baseball players ran from home plate to second base two times for each of three routes: *round*, *narrow*, and *wide*. These routes are illustrated in the figure below (the routes have been exaggerated slightly for illustration). Let's compare the *narrow* and *wide* routes. (Note: The running times are the average of two runs between a point 35 feet from home plate to a point 15 feet short of second base). For the sample of observations, the mean difference is $\bar{x} = 0.075$ seconds, and the standard deviation is $s = 0.088$ seconds. Is a mean difference of 0.075 seconds statistically significant at a significance level of $\alpha = 0.01$?

	Route		
Player	**narrow**	**wide**	**Difference**
1	5.5	5.55	-0.05
2	5.7	5.75	-0.05
3	5.6	5.5	0.1
4	5.5	5.4	0.1
5	5.85	5.7	0.15
⋮	⋮	⋮	⋮
22	6.3	6.25	0.05



Types of errors in significance tests

- There are two types of wrong decisions in significance testing:
- **Type I error** – When H_0 is true, but we reject the null hypothesis (false positive result)
- **Type II error –** When H_0 is false but we fail to reject the null hypothesis (false negative result)

	Decision	
Reality	**Do Not Reject H_0**	**Reject H_0**
H_0 true	correct decision	type I error
H_0 false	type II error	correct decision

Ex.) Types of errors in significance tests

- Recall the twin study that examined the relationship between schizophrenia and left hippocampus volume. Suppose the hypotheses are $H_0\colon \mu_0 = 0$ (there is no relationship) and $H_A\colon \mu \neq \mu_0$ (there is a relationship).

	Decision	
Reality	**Do Not Reject H_0**	**Reject H_0**
there is no relationship	correctly conclude there is no relationship	incorrectly conclude there is a relationship
there is a relationship	incorrectly conclude there is no relationship	correctly conclude there is a relationship

- We rejected H_0, if H_0 is true, what kind of error did we make?

Ex.) Types of errors in significance tests

- Recall the study with the cross-over design that investigated if garlic repels ticks. Suppose the hypotheses are $H_0: p_0 = 0.5$ (garlic is not effective) versus $H_A: p > p_0$ (garlic is effective).

	Decision	
Reality	**Do Not Reject H_0**	**Reject H_0**
garlic is not effective	correctly conclude that garlic is ineffective	incorrectly conclude that garlic is effective
garlic is effective	incorrectly conclude that garlic is ineffective	correctly conclude that garlic is effectve

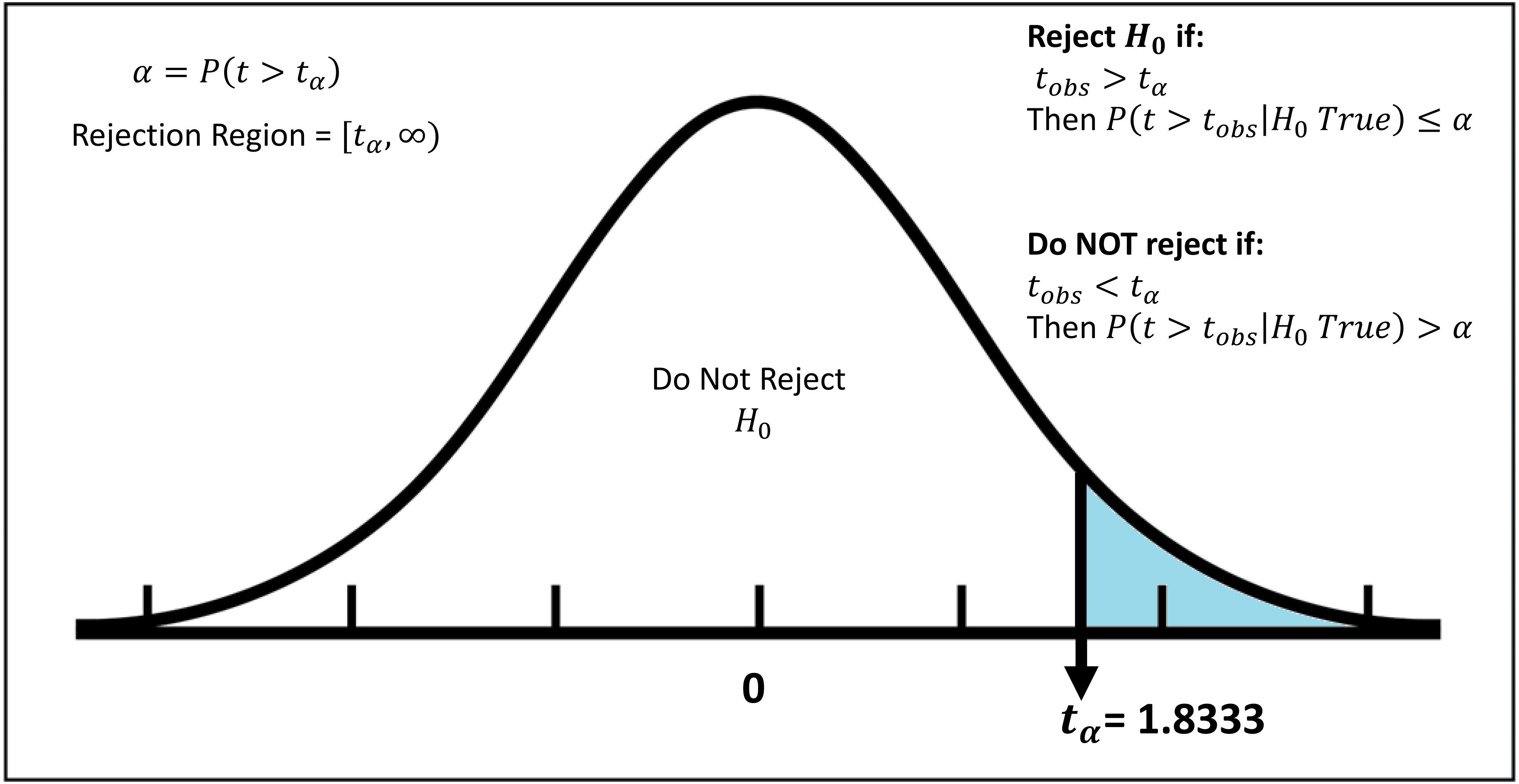
- We failed to reject H_0, if H_0 is actually false, what kind of error did we make?

Probability of a type I error

- The probability of committing a type I error is the probability of rejecting H_0 when H_0 is true:

- Suppose we have the hypotheses $H_0\colon \mu_0 = 0$ versus $H_A\colon \mu > \mu_0$ and plan to use a significance level of α=0.05. The critical value $t_\alpha = 1.833$ is the value of the test statistic with a P-value equal to the significance level. Assume a sample size of n=10.

Suppose we have the hypotheses $H_0: \mu_0 = 0$ versus $H_A: \mu > \mu_0$ and plan to use a significance level of $\alpha = 0.05$. The critical value t_α is the value of the test statistic with a P-value equal to the significance level. Assume a sample size of $n = 10$.



Probability of a type I error

- The probability of a type I error is the probability of rejecting H_0 when H_0 is true

$$P(t \geq t_\alpha | H_0\, True) = \alpha$$
$$P(t \geq 1.8333 | H_0\, True) = 0.05$$

Thus, the probability of committing a type I error is α

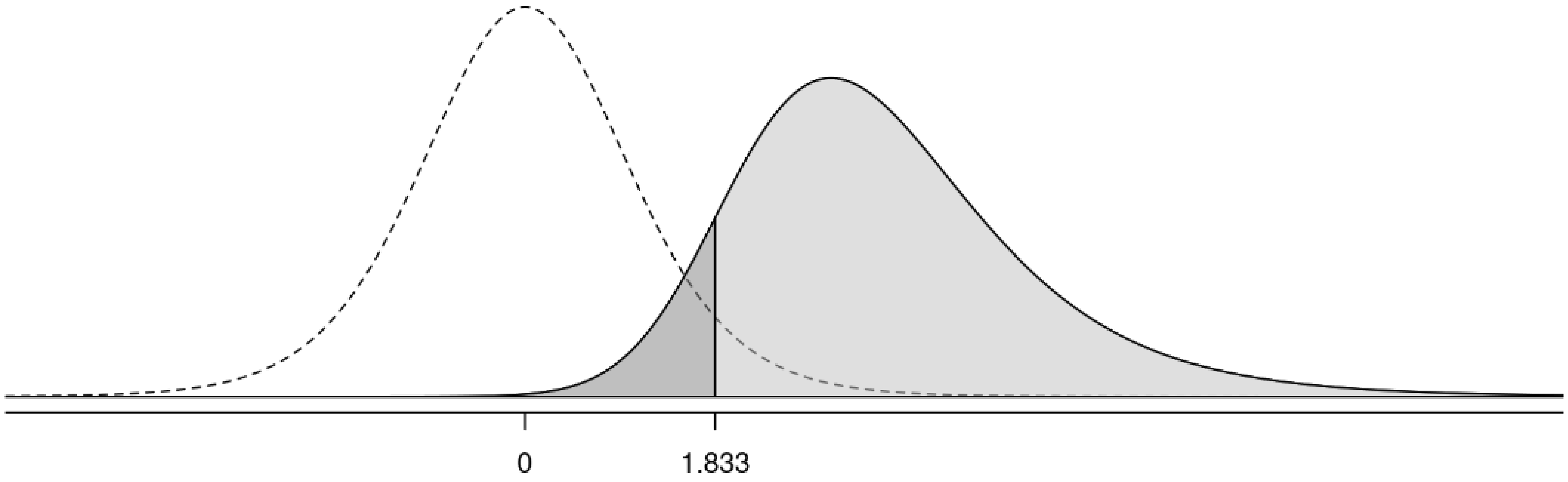
Probability of a type II error

- The probability of a type II error (denoted β) is the probability of not rejecting H_0 when it is false.

- Suppose again we have the hypotheses $H_0: \mu_0 = 0$ versus $H_A: \mu > \mu_0$ and plan to use a significance level of α=0.05. The critical value of t is the value of the test statistic with a P-value equal to the significance level. Assume a sample size of $n = 10$. But now suppose that in reality $\mu > 0$ (e.g., $\mu = 1$). Note that the sampling distribution of the test statistic when H_0 is true is shown by the dotted line, while the sampling distribution of the test statistic when H_0 is false is shown by the solid line.

Probability of a type II error

- But now suppose that in reality $\mu > 0$ (e.g., $\mu = 1$). Note that the sampling distribution of the test statistic when H_0 is true is shown by the dotted line, while the sampling distribution of the test statistic when H_0 is false is shown by the solid line.



Probability of a type II error

- So, the probability of a type II error (i.e., the probability of not rejecting H_0 when it is false) here is

$$\beta = P(t < t_\alpha | H_A\ True)$$

$$\beta = P(t < 1.8333\ | H_A\ True)$$

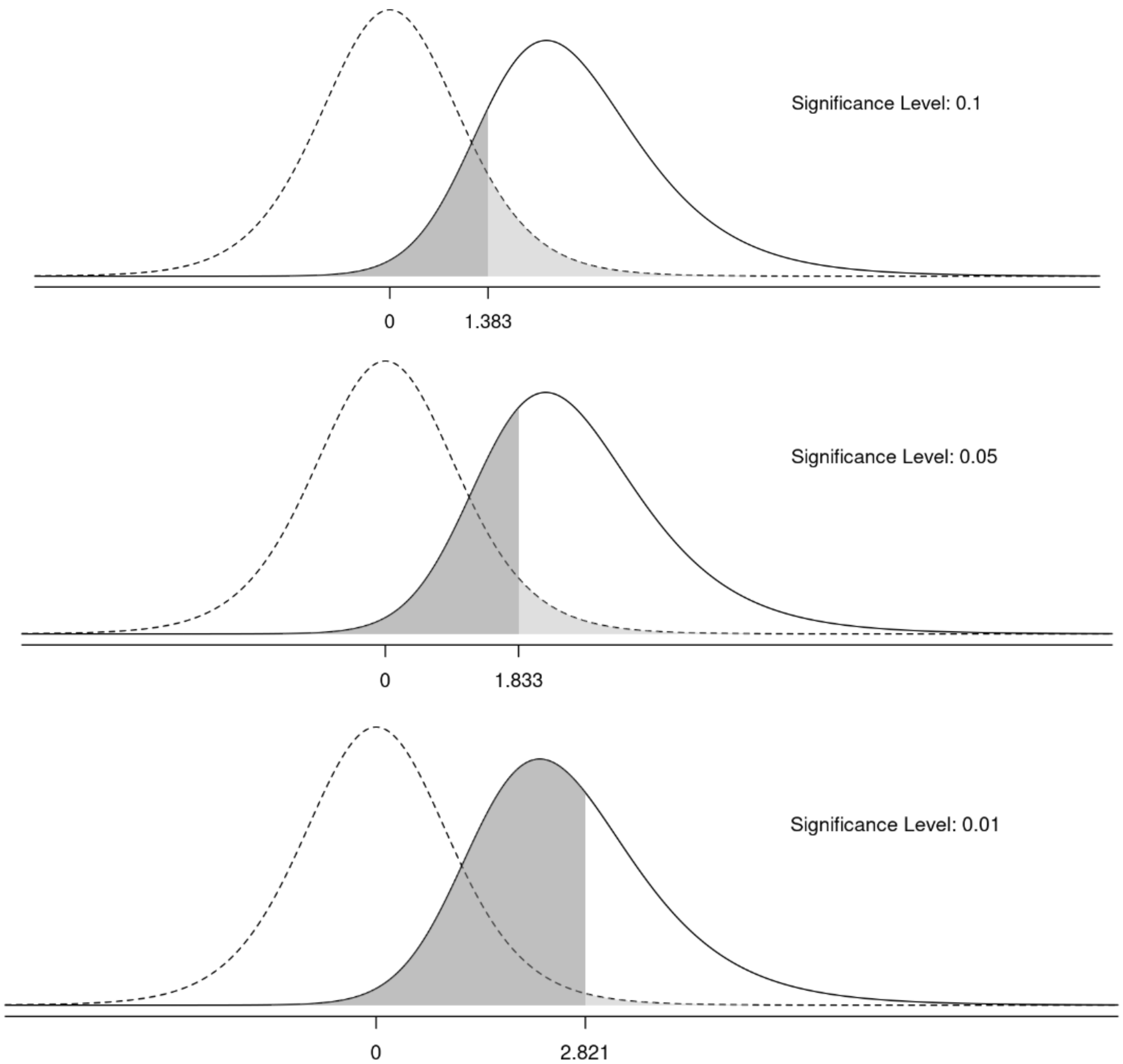
- It is not as simple to compute the probability of a type II error because it depends on several factors

The effect of α on error probabilities

- Decreasing the significance level α (a.) decreases the type I error rate but (b.) increases the type II error rate

- Increasing the significance level α (a.) increases the type I error rate but (b.) decreases the type II error rate

The probability of a type I error is the *light* grey area, and the probability of a type II error is the *dark* grey area.



Power

- The statistical **power** of a significance test is the probability of rejecting the null hypothesis when it is false

$$Power = P(t > t_{\alpha} | \mathrm{H_A\ true}) = 1 - P(t < t_{\alpha} | \mathrm{H_A\ true}) = 1 - \beta$$

- It is the probability that we don't commit a type II error

Ways to increase power:

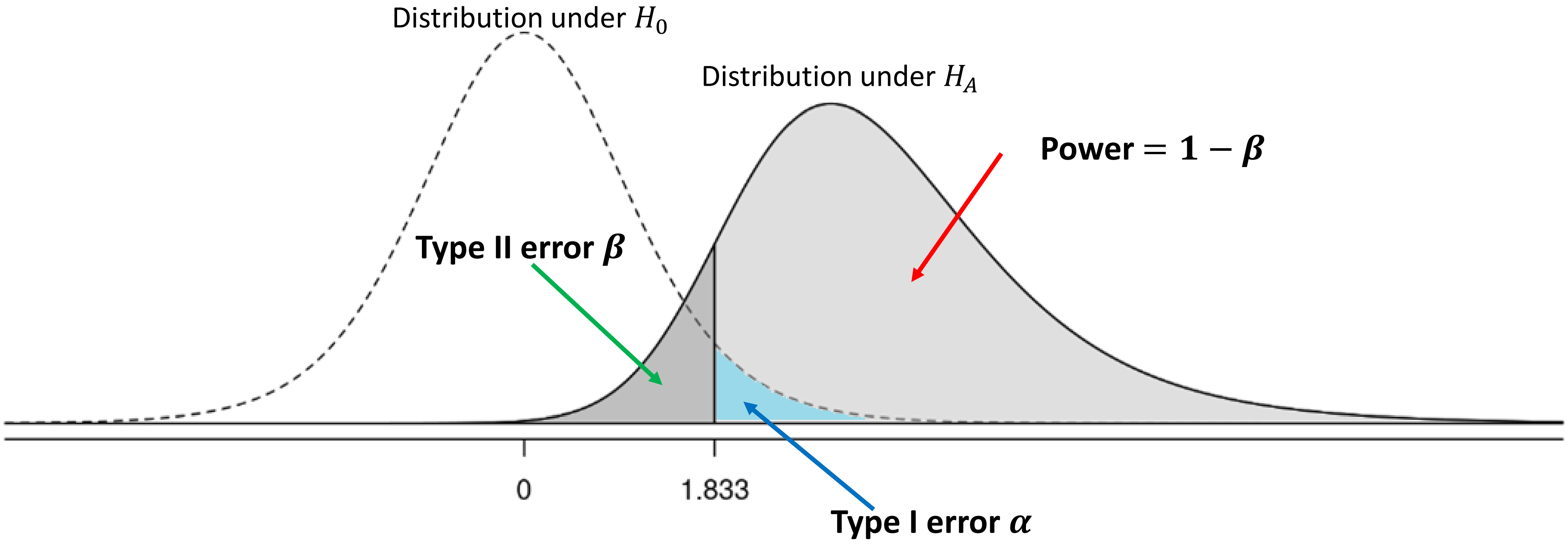
- Increase the sample size n
- Increase the significance level α
- Choose a more powerful test

Significance Level (α) and Power (1-β)

	H_0 is True	H_0 is False
Test Rejects H_0	α	1-β
Test Doesn't Reject H_0	1-α	β

Type I, Type II, and Power

- Using our previous examples for type I and II error. We have hypotheses $H_0: \mu_0 = 0$ versus $H_A: \mu > \mu_0$ and plan to use a significance level of α=0.05. The critical value of t is the value of the test statistic with a P-value equal to the significance level. Assume a sample size of $n = 10$.
- But now suppose that in reality $\mu > 0$ (e.g., $\mu = 1$). Note that the sampling distribution of the test statistic when H_0 is true is shown by the dotted line, while the sampling distribution of the test statistic when H_0 is false is shown by the solid line.



Relationship between **power**, α, and β