INTRODUCTION TO STATISTICS

LECTURE 11

LAST TIME

- Ingredients for hypothesis testing:
  - H0 and H1
  - Test statistic
  - Null distribution
  - Rejection region
  - Two- and one-sided alternatives

LAST TIME

- Ingredients for hypothesis testing:
  - H0 and H1
  - Test statistic
  - Null distribution
  - Rejection region
  - Two- and one-sided alternatives

- One-sample tests
  - z-test
  - t-test

TODAY

- Types of errors
- Power of the test
- p-values
- More tests
- Practice!

HYPOTHESIS TESTING STEP-BY-STEP

- Collect data X

HYPOTHESIS TESTING STEP-BY-STEP

- Collect data X

- Set up H_0 and H_1
  - two-sided or one-sided

HYPOTHESIS TESTING STEP-BY-STEP

- Collect data X
- Set up H_0 and H_1
  - two-sided or one-sided
- Chose test statistic $T(X)$
- Determine distribution of T assuming H_0

HYPOTHESIS TESTING STEP-BY-STEP

- Collect data X
- Set up H_0 and H_1
  - two-sided or one-sided
- Chose test statistic $T(X)$
- Determine distribution of T assuming H_0
- Determine rejection area R

HYPOTHESIS TESTING STEP-BY-STEP

- Collect data X
- Set up H_0 and H_1
  - two-sided or one-sided
- Chose test statistic $T(X)$
- Determine distribution of T assuming H_0
- Determine rejection area R
- Compute the value t of $T(X)$ from data

HYPOTHESIS TESTING STEP-BY-STEP

- Collect data X
- Set up H_0 and H_1
    - two-sided or one-sided
- Chose test statistic $T(X)$
- Determine distribution of T assuming H_0
- Determine rejection area R
- Compute the value t of $T(X)$ from data
- Check if it falls into the rejection area R
    - YES $\Longrightarrow$ reject H_0
    - NO $\Longrightarrow$ don't reject H_0

ONE-SAMPLE TESTS

Check if the mean of the data equals some hypothesized value assuming that the data is normal

Z-TEST

- Data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$$

- Mean μ is unknown, and the variance σ^2 is **known.**

Z-TEST

- Data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$$

- Mean μ is unknown, and the variance σ^2 is **known.**
- Null hypothesis: H_0: $\mu_X = \mu_0$

Z-TEST

- Data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$$

- Mean μ is unknown, and the variance σ^2 is **known.**
- Null hypothesis: H_0: $\mu_X = \mu_0$
- Alternatives:
  - Two-sided: $H_1 : \mu_X \neq \mu_0$
  - One-sided: $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$

Z-TEST

- Test statistic:

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$$

Z-TEST

- Test statistic:

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}$$

- Null distribution:

$$\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$$

Z-TEST: EXAMPLE

- Assume that weight loss follows $N(\mu, 7^2)$. 30 patients followed an experimental diet. Average weight loss $\bar{X} = 6.1$ kg. Is there evidence that $\mu \neq 5$? Test at $\alpha = 0.05$.

Z-TEST: EXAMPLE

- Assume that weight loss follows $N(\mu, 7^2)$. 30 patients followed an experimental diet. Average weight loss $\bar{X} = 6.1$ kg. Is there evidence that $\mu \neq 5$? Test at $\alpha = 0.05$.

$$H_0: \mu = 5, \qquad H_1: \mu \neq 5$$

Z-TEST: EXAMPLE

- Assume that weight loss follows $N(\mu, 7^2)$. 30 patients followed an experimental diet. Average weight loss $\bar{X} = 6.1$ kg. Is there evidence that $\mu \neq 5$? Test at $\alpha = 0.05$.

$$H_0: \mu = 5, \qquad H_1: \mu \neq 5$$

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1), \qquad z_{1-0.025} = 1.96$$

Z-TEST: EXAMPLE

- Assume that weight loss follows $N(\mu, 7^2)$. 30 patients followed an experimental diet. Average weight loss $\bar{X} = 6.1$ kg. Is there evidence that $\mu \neq 5$? Test at $\alpha = 0.05$.

$$H_0\colon \mu = 5, \qquad H_1\colon \mu \neq 5$$

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1), \qquad z_{1-0.025} = 1.96$$

$$T(X) = \frac{(6.1-5)\sqrt{30}}{7} = 0.86$$

Z-TEST: EXAMPLE

- Assume that weight loss follows $N(\mu, 7^2)$. 30 patients followed an experimental diet. Average weight loss $\bar{X} = 6.1$ kg. Is there evidence that $\mu \neq 5$? Test at $\alpha = 0.05$.

$$H_0: \mu = 5, \qquad H_1: \mu \neq 5$$

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1), \qquad z_{1-0.025} = 1.96$$

$$T(X) = \frac{(6.1-5)\sqrt{30}}{7} = 0.86, \quad |T(X)| < 1.96 \implies \text{don't reject } H_0$$

T-TEST

- Data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$$

- Mean μ and the variance σ^2 are **unknown.**

T-TEST

- Data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$$

- Mean μ and the variance σ^2 are **unknown.**
- Null hypothesis: $H_0\colon \mu_X = \mu_0$

T-TEST

- Data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$$

- Mean μ and the variance σ^2 are **unknown.**
- Null hypothesis: H_0: $\mu_X = \mu_0$
- Alternatives:
  - Two-sided: $H_1: \mu_X \neq \mu_0$
  - One-sided: $H_1: \mu > \mu_0$ or $H_1: \mu < \mu_0$

T-TEST

- Test statistic:

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{X} - X_i)^2$$

T-TEST

- Test statistic:

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{X} - X_i)^2$$

- Null distribution:

$$\frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim t(n-1)$$

T-TEST: EXAMPLE

- Based of a sample of $\mathrm{n} = 25$ salesmen, average sale is $\bar{X} = 130$ dollars, with sample standard deviation $\mathrm{s} = 15$. Are the sales larger than \$100 on average? Test at $\alpha = 0.05$. Assume that sales follow $N(\mu, \sigma^2)$.

T-TEST: EXAMPLE

- Based of a sample of $\mathrm{n} = 25$ salesmen, average sale is $\bar{X} = 130$ dollars, with sample standard deviation $\mathrm{s} = 15$. Are the sales larger than \$100 on average? Test at $\alpha = 0.05$. Assume that sales follow $N(\mu, \sigma^2)$.

$$H_0: \mu = 100, \qquad H_1: \mu > 100$$

T-TEST: EXAMPLE

- Based of a sample of $\mathrm{n} = 25$ salesmen, average sale is $\bar{X} = 130$ dollars, with sample standard deviation $\mathrm{s} = 15$. Are the sales larger than $100 on average? Test at $\alpha = 0.05$. Assume that sales follow $N(\mu, \sigma^2)$.

$$H_0: \mu = 100, \qquad H_1: \mu > 100$$

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim t(24),$$

T-TEST: EXAMPLE

- Based of a sample of $\mathrm{n} = 25$ salesmen, average sale is $\bar{X} = 130$ dollars, with sample standard deviation $\mathrm{s} = 15$. Are the sales larger than $100 on average? Test at $\alpha = 0.05$. Assume that sales follow $N(\mu, \sigma^2)$.

$$H_0: \mu = 100, \qquad H_1: \mu > 100$$

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim t(24), \qquad t_{1-0.05}(24) = 1.711$$

T-TEST: EXAMPLE

- Based of a sample of $\mathrm{n} = 25$ salesmen, average sale is $\bar{X} = 130$ dollars, with sample standard deviation $\mathrm{s} = 15$. Are the sales larger than \$100 on average? Test at $\alpha = 0.05$. Assume that sales follow $N(\mu, \sigma^2)$.

$$H_0: \mu = 100, \qquad H_1: \mu > 100$$

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim t(24), \qquad t_{1-0.05}(24) = 1.711$$

$$T(X) = \frac{(130-100)\sqrt{25}}{15} = 10$$

T-TEST: EXAMPLE

- Based of a sample of $\mathrm{n} = 25$ salesmen, average sale is $\bar{X} = 130$ dollars, with sample standard deviation $\mathrm{s} = 15$. Are the sales larger than \$100 on average? Test at $\alpha = 0.05$. Assume that sales follow $N(\mu, \sigma^2)$.

$$H_0: \mu = 100, \qquad H_1: \mu > 100$$

$$T(X) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim t(24), \qquad t_{1-0.05}(24) = 1.711$$

$$T(X) = \frac{(130-100)\sqrt{25}}{15} = 10 > 1.711 \implies \text{reject } H_0$$

TYPES OF ERRORS

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0		
	Don't reject H_0		

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0		correct decision
	Don't reject H_0	correct decision	

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	Don't reject H_0	correct decision	Type II error

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	Don't reject H_0	correct decision	Type II error

- Significance level:

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	Don't reject H_0	correct decision	Type II error

- Significance level:

$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0) = P(T \in R \mid H_0)$$

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	Don't reject H_0	correct decision	Type II error

- Significance level:

$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0) = P(T \in R \mid H_0)$$

- Power:

$$P(\text{Reject } H_0 \mid H_1) = P(T \in R \mid H_1) =$$

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	Don't reject H_0	correct decision	Type II error

- Significance level:

$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0) = P(T \in R \mid H_0)$$

- Power:

$$P(\text{Reject } H_0 \mid H_1) = P(T \in R \mid H_1) = 1 - P(\text{Type II error})$$

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	Don't reject H_0	correct decision	Type II error

- Significance level:

$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0) = P(T \in R \mid H_0)$$

- Power:

$$P(\text{Reject } H_0 \mid H_1) = P(T \in R \mid H_1) = 1 - P(\text{Type II error})$$

We want power near and significance near

TYPES OF ERRORS

		True state of nature	
		H_0	H_A
Our decision	Reject H_0	Type I error	correct decision
	Don't reject H_0	correct decision	Type II error

- Significance level:

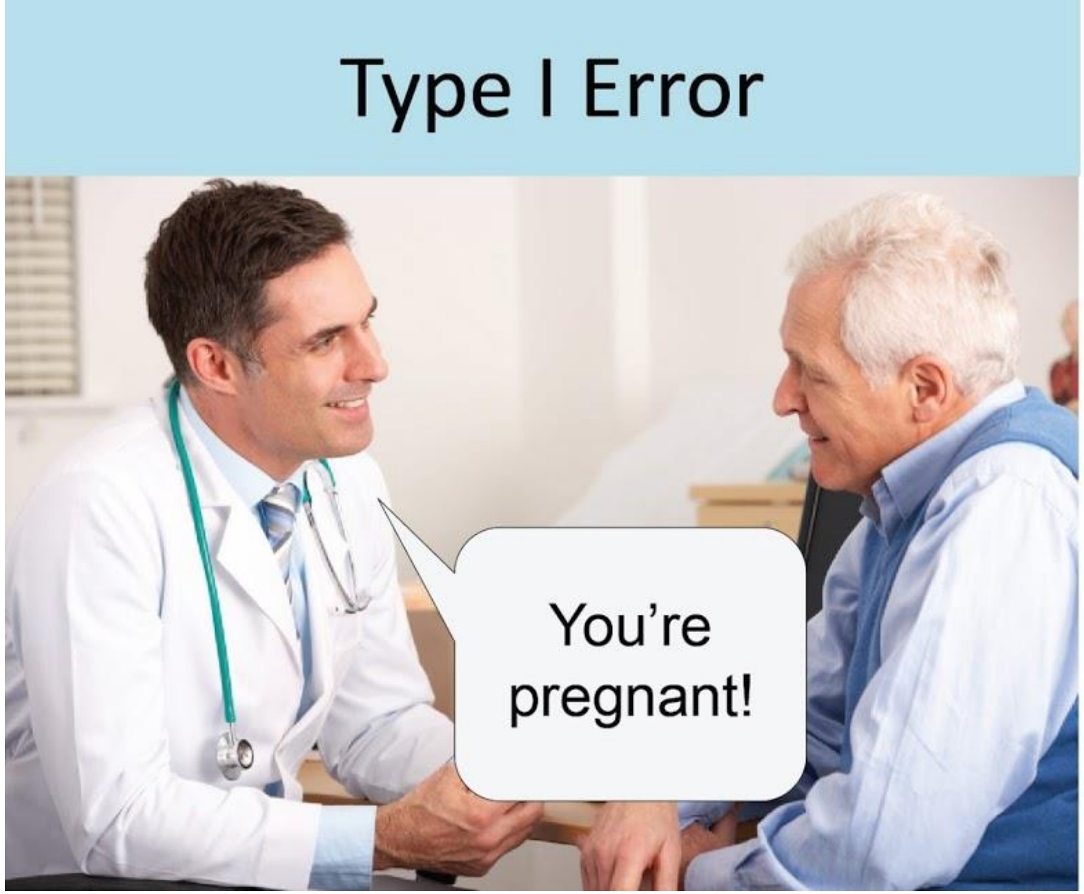
$$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0) = P(T \in R \mid H_0)$$

- Power:

$$P(\text{Reject } H_0 \mid H_1) = P(T \in R \mid H_1) = 1 - P(\text{Type II error})$$

We want power near 1 and significance near 0

TYPES OF ERRORS



P-VALUES

P-VALUES

- Up till now:
  - Rejecting H_0 based on the value of the test statistic:

$$T \in R \Longrightarrow \text{reject } H_0, \qquad T \notin R \Longrightarrow \text{do not reject } H_0$$

P-VALUES

- Up till now:
  - Rejecting H_0 based on the value of the test statistic:

$$T \in R \implies \text{reject } H_0, \qquad T \notin R \implies \text{do not reject } H_0$$

- Not very practical:
  - The value of the test statistic alone isn't informative.
  - Rejection region R is different for every test.

P-VALUES

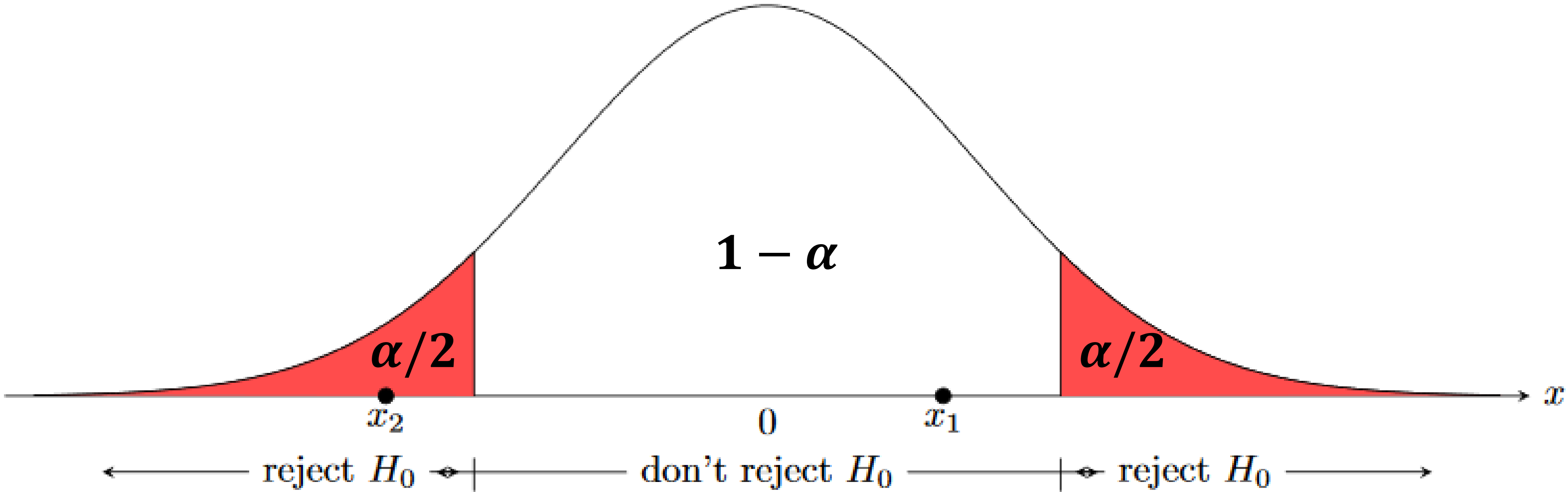
- Up till now:
  - Rejecting H_0 based on the value of the test statistic:

$$T \in R \Longrightarrow \text{reject } H_0, \qquad T \notin R \Longrightarrow \text{do not reject } H_0$$

- Not very practical:
  - The value of the test statistic alone isn't informative.
  - Rejection region R is different for every test.

- A more unified approach: report **p-value**.

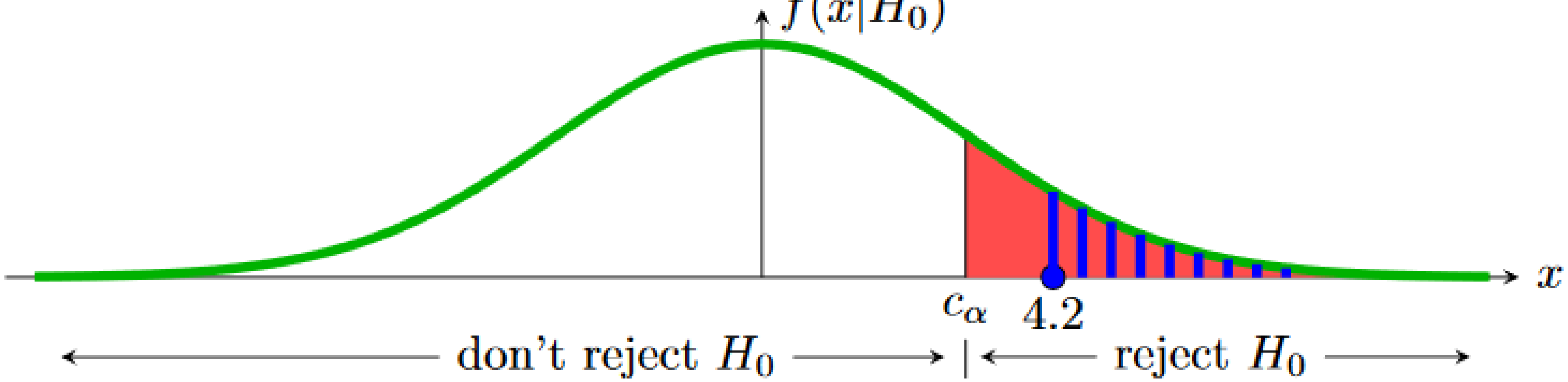
P-VALUES



$1-\alpha$
$\alpha/2$
$\alpha/2$
x
x_2
0
x_1
reject H_0
don't reject H_0
reject H_0

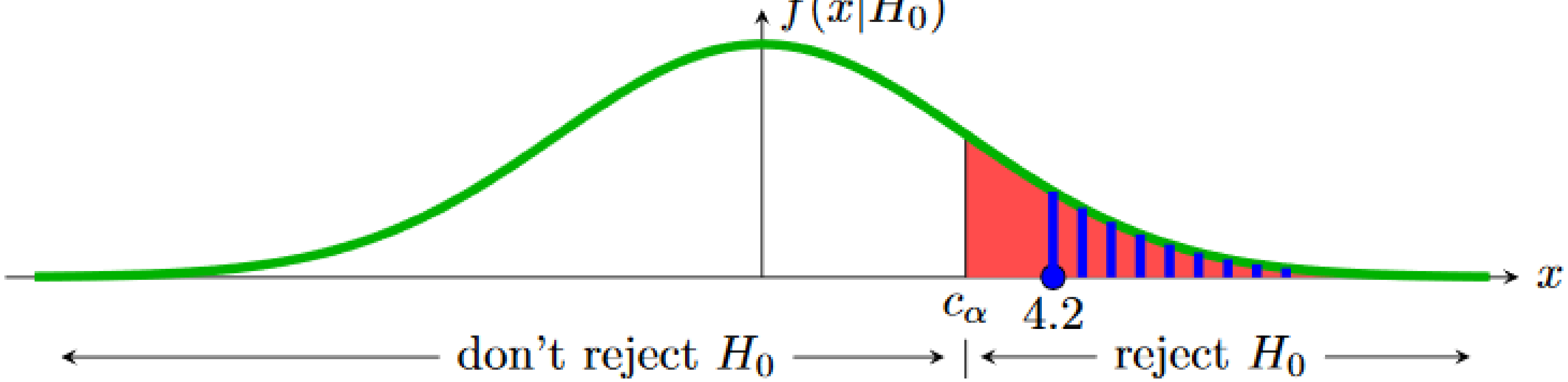
P-VALUES

- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t = 4.2. Should we reject H_0?



P-VALUES

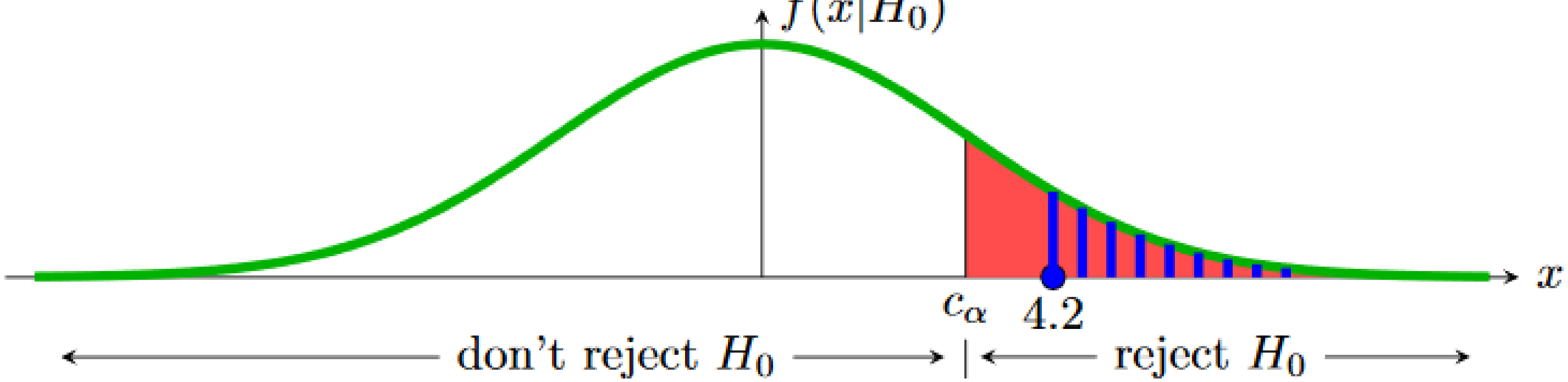
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t $=$ 4.2. Should we reject H_0?



- Answer:
  - Yes, if $T = 4.2$ is in the rejection region.

P-VALUES

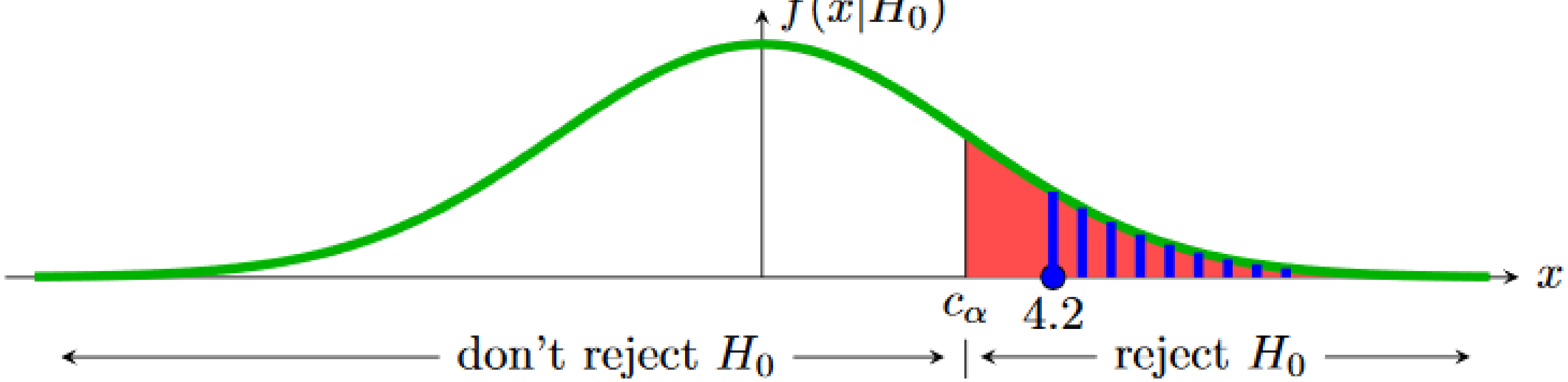
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t = 4.2. Should we reject H_0?



- Answer:
  - Yes, if $T = 4.2$ is in the rejection region.
- Alternatively:
  - Yes, if **blue area** < **red area**

P-VALUES

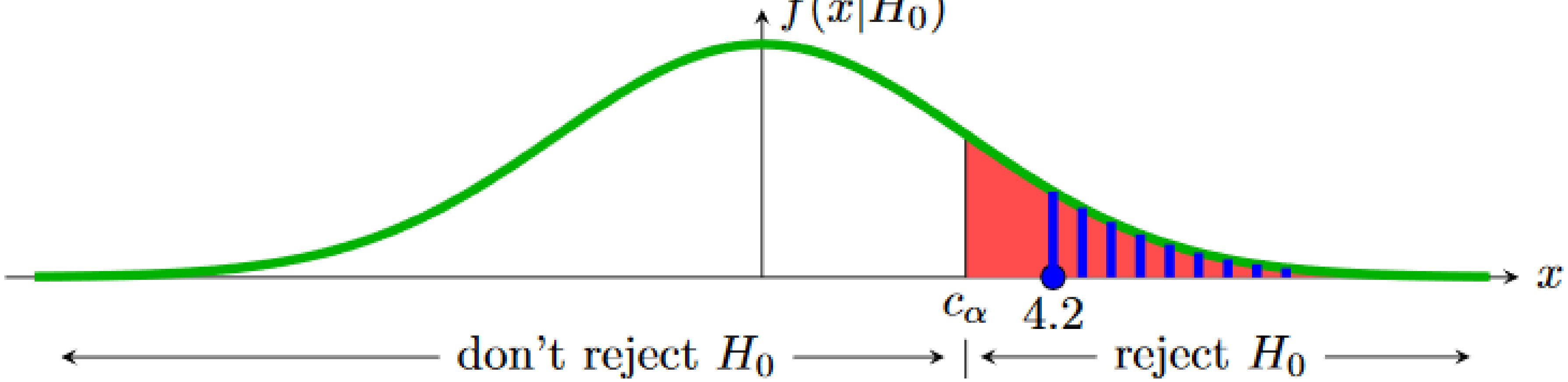
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t $=$ 4.2. Should we reject H_0?



- Significance: $\alpha \; = \; P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area

P-VALUES

- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t = 4.2. Should we reject H_0?

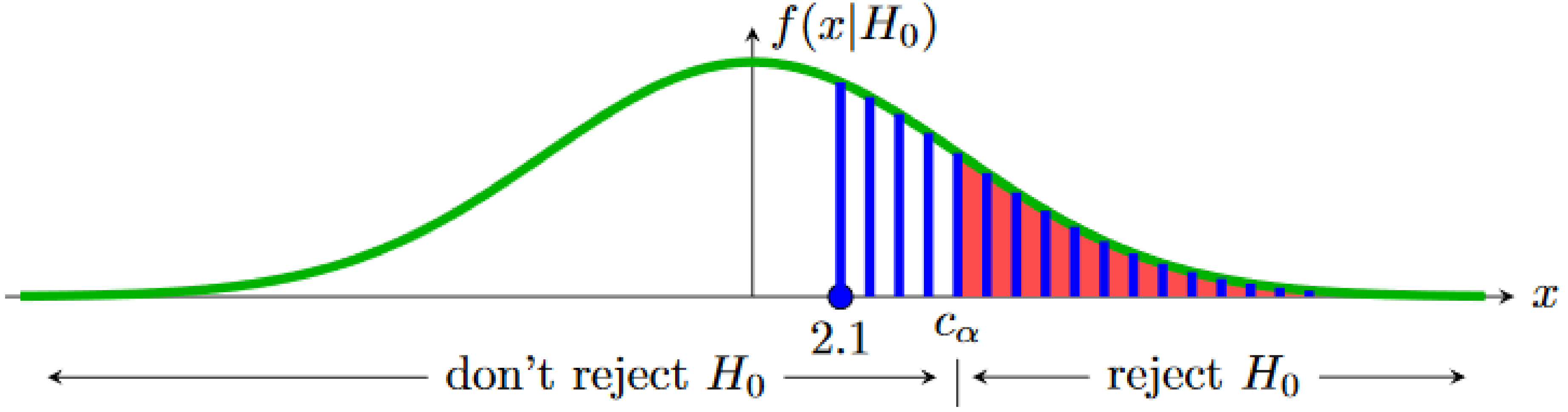


- Significance: $\alpha = P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area

Since $p < \alpha$, reject H_0

P-VALUES

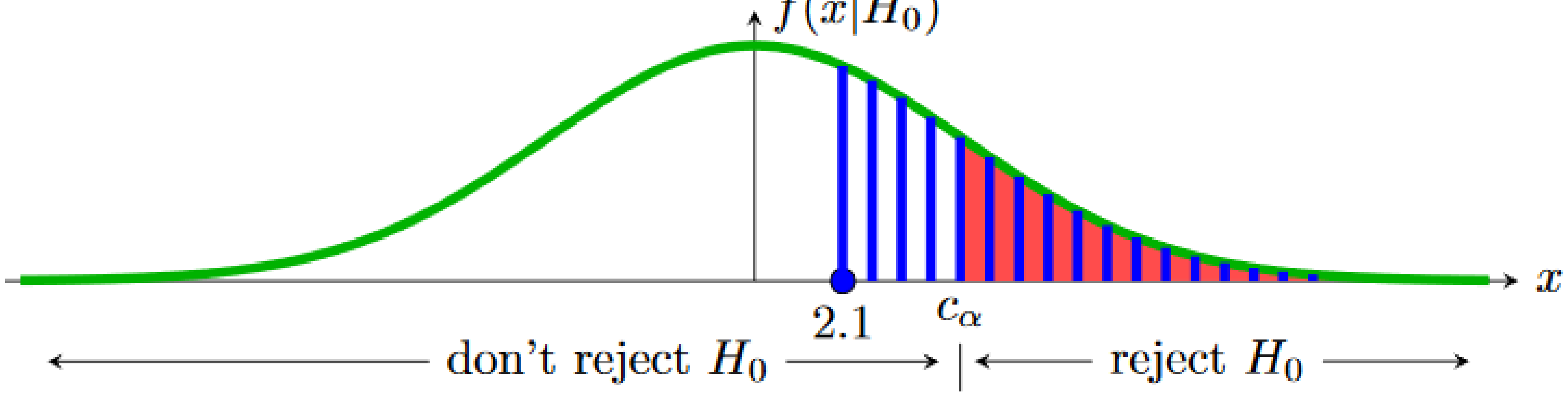
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t $= 2.1$. Should we reject H_0?



- Significance: $\alpha = P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area

P-VALUES

- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t $= 2.1$. Should we reject H_0?

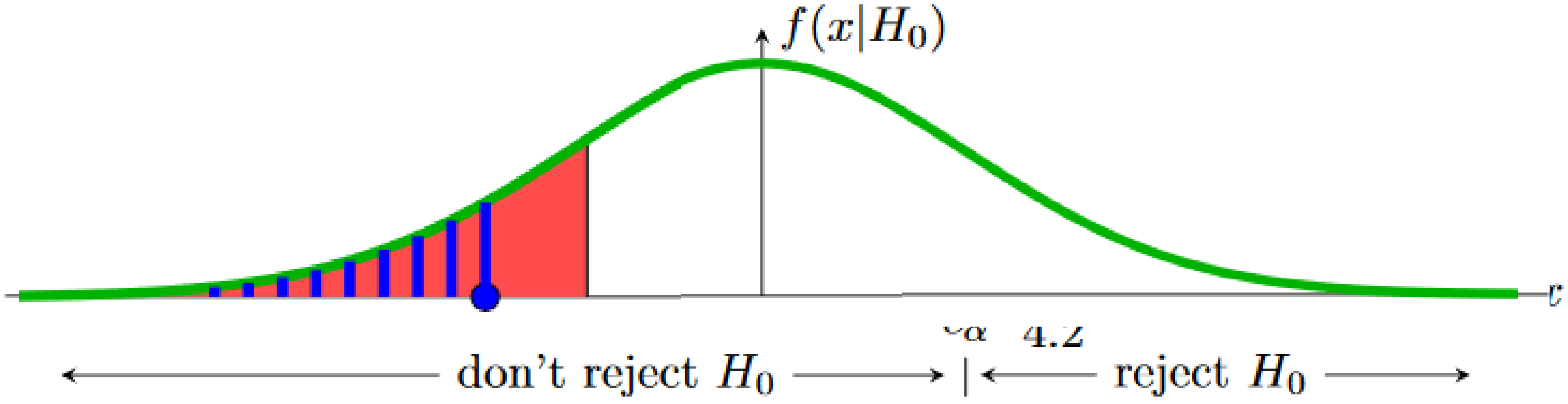


- Significance: $\alpha \;=\; P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area

Since $p > \alpha$, do not reject H_0

P-VALUES

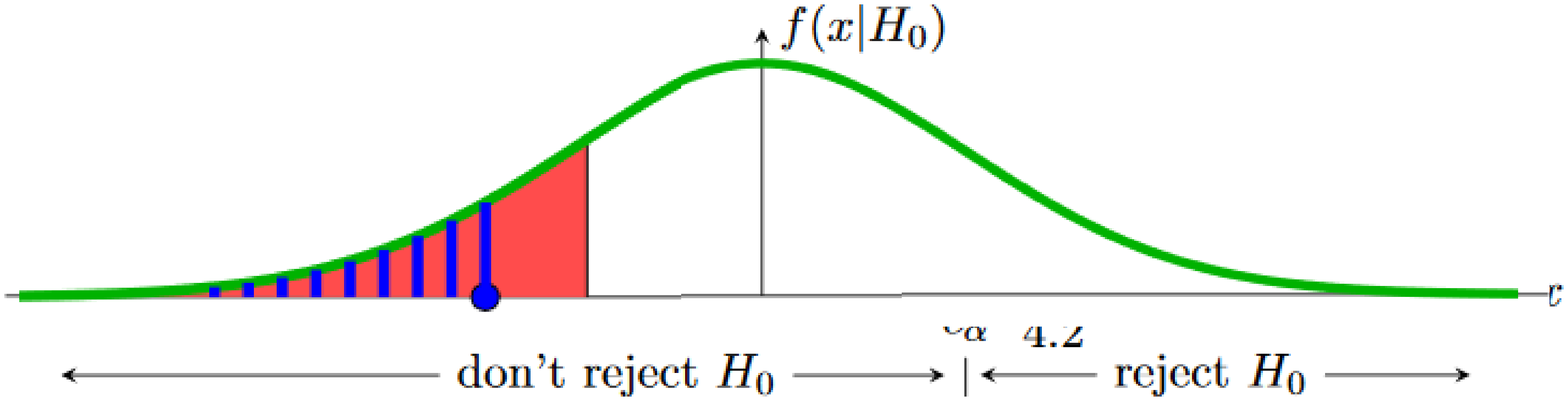
- Similarly, with the other one-sided alternative:
- Significance: $\alpha = P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area



P-VALUES

- Similarly, with the other one-sided alternative:
- Significance: $\alpha \;=\; P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area

Since $p < \alpha$, reject H_0

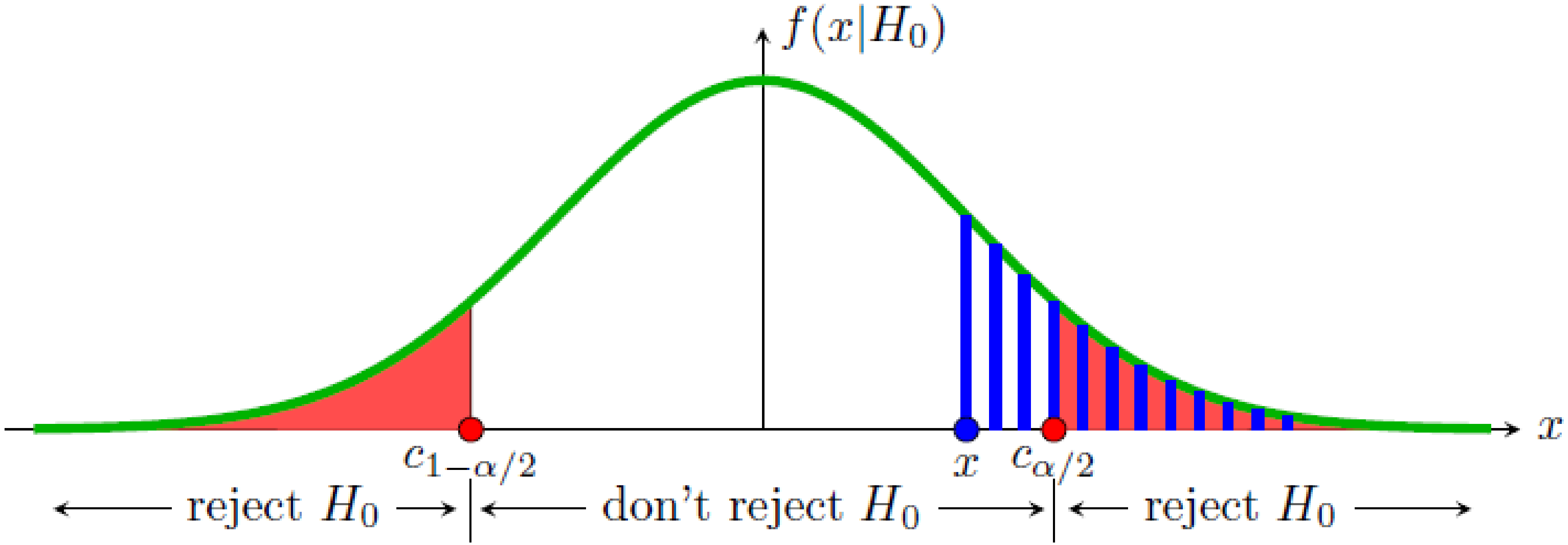


P-VALUES

- Two-sided p-values are a bit trickier:
  - What does ‘data at least as extreme’ means in this context?

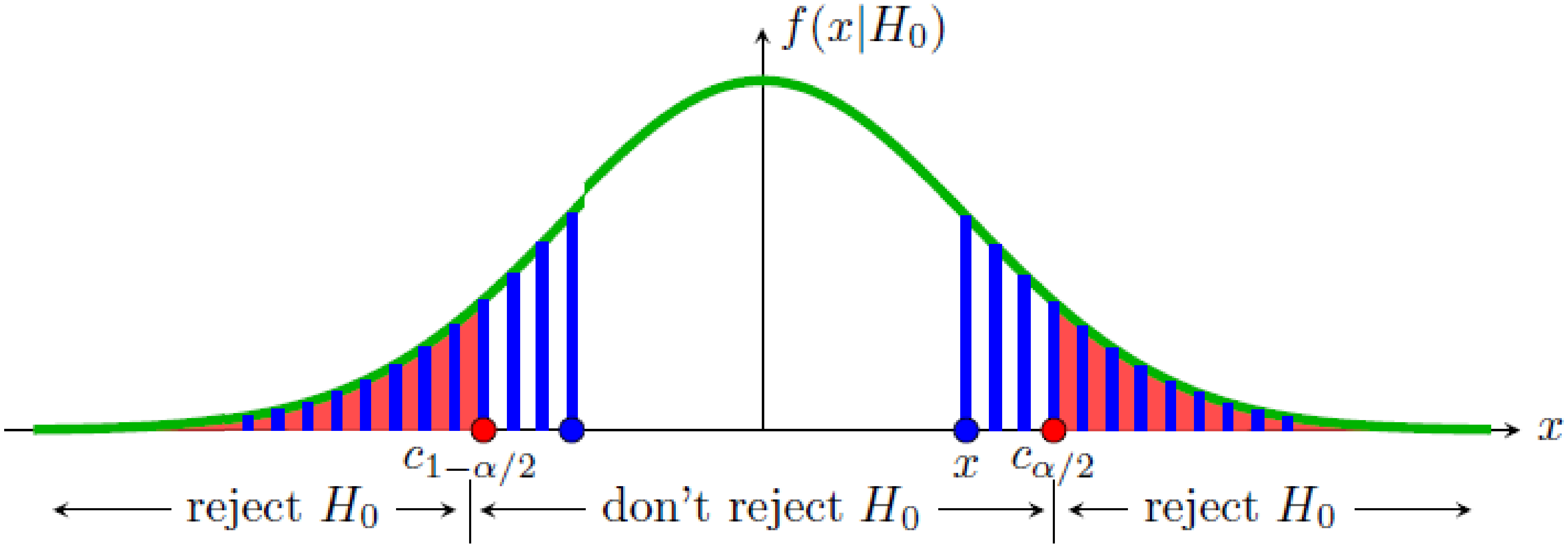
P-VALUES

- Two-sided p-values are a bit trickier:
  - What does ‘data at least as extreme’ means in this context?



P-VALUES

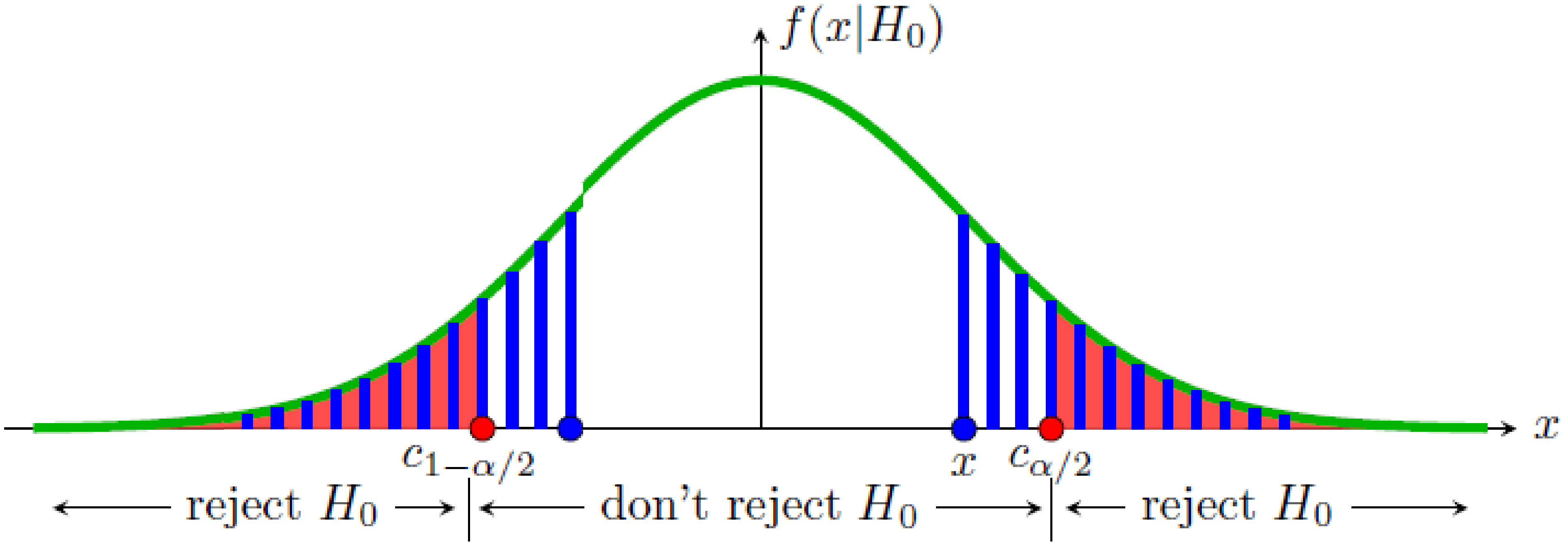
- Two-sided p-values are a bit trickier:
  - What does 'data at least as extreme' means in this context?



P-VALUES

- Two-sided p-values are a bit trickier:
  - What does ‘data at least as extreme’ means in this context?

$$p = \mathbf{2}\min\{P(T > t), P(T \leq t)\}$$



P-VALUES

- Summing it up:

if t is the value of test statistic, $p = \cdots$

P-VALUES

- Summing it up:

if t is the value of test statistic, $p = \cdots$

  - One-sided alternative ('greater,' right tail): $p = P(T > t)$

P-VALUES

- Summing it up:

if t is the value of test statistic, $p = \cdots$

  - One-sided alternative ('greater,' right tail): $p = P(T > t)$
  - One-sided alternative ('smaller', left tail): $p = P(T \leq t)$

P-VALUES

- Summing it up:

if t is the value of test statistic, $p = \cdots$

  - One-sided alternative (‘greater,’ right tail): $p = P(T > t)$
  - One-sided alternative (‘smaller’, left tail): $p = P(T \leq t)$
  - Two-sided alternative: $p = 2\min\{P(T > t), P(T \leq t)\}$

P-VALUES

- Summing it up:

if t is the value of test statistic, $p = \cdots$

- One-sided alternative ('greater,' right tail): $p = P(T > t)$
- One-sided alternative ('smaller', left tail): $p = P(T \le t)$
- Two-sided alternative: $p = 2\min\{P(T > t), P(T \le t)\}$

$\boldsymbol{p < \alpha \implies}$ **Reject** $\boldsymbol{H_0}$

PRACTICE!

Google Classroom -> Lecture 11 -> One-sample tests

MORE TESTS

MORE TESTS

- There exist a lot of tests.
- Don’t try to memorize all the tests – there’re too many of them.
- Your task is to find the right test when you need it.
- Here are some of the most used ones **(but not all)**.

TWO-SAMPLE TESTS

TWO-SAMPLE TESTS

- Consider the task of comparing the means of two **independent** samples.

TWO-SAMPLE TESTS

- Consider the task of comparing the means of two **independent** samples.

- Examples:
  - comparing the mean efficacies of two medical treatments;
  - comparing academic performance in two different groups;

TWO-SAMPLE T-TESTS

Test difference in the means of two populations
with unknown but equal variances

TWO-SAMPLE T-TEST

- Two sets of data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \ldots, X_n \sim N(\mu_X, \sigma^2)$$

$$Y_1, Y_2, \ldots, Y_m \sim N(\mu_Y, \sigma^2)$$

- Means μ_X and μ_Y and the variance σ^2 are **unknown.**
  - Note that the two distributions have **the same variance**.

TWO-SAMPLE T-TEST

- Two sets of data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \dots, X_n \sim N(\mu_X, \sigma^2)$$

$$Y_1, Y_2, \dots, Y_m \sim N(\mu_Y, \sigma^2)$$

- Means μ_X and μ_Y and the variance σ^2 are **unknown.**
  - Note that the two distributions have **the same variance**.
- Null hypothesis: H_0: $\mu_X - \mu_Y = 0$

TWO-SAMPLE T-TEST

- Two sets of data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \dots, X_n \sim N(\mu_X, \sigma^2)$$

$$Y_1, Y_2, \dots, Y_m \sim N(\mu_Y, \sigma^2)$$

- Means μ_X and μ_Y and the variance σ^2 are **unknown.**
  - Note that the two distributions have **the same variance**.
- Null hypothesis: H_0: $\mu_X - \mu_Y = 0$
- Alternatives:
  - Two-sided: H_1: $\mu_X - \mu_Y \neq 0$
  - One-sided: H_1: $\mu_X - \mu_Y > 0$ or H_1: $\mu_X - \mu_Y < 0$

TWO-SAMPLE T-TEST

- Test statistic:

$$T(X,Y) = \frac{\bar{X} - \bar{Y}}{s_p}\text{, where}$$

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} \cdot \left(\frac{1}{n} + \frac{1}{m}\right)$$

s_X^2, s_Y^2 - sample variances of X and Y.

TWO-SAMPLE T-TEST

- Test statistic:

$$T(X,Y) = \frac{\bar{X}-\bar{Y}}{s_p}\text{, where}$$

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} \cdot \left(\frac{1}{n} + \frac{1}{m}\right)$$

s_X^2, s_Y^2 - sample variances of X and Y.

- Null distribution:

$$\frac{\bar{X}-\bar{Y}}{s_p} \sim t(n+m-2)$$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

Set up and run two-sample t-test to check if the mean duration of pregnancy differs in the two groups.

Which assumptions do you have to make?

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

H_0:

H_1:

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

$$H_0\colon \; \mu_x - \mu_Y = 0$$

$$H_1\colon \; \mu_x - \mu_Y \neq 0$$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

$$T = \frac{\bar{X} - \bar{Y}}{s_p} \sim t(n + m - 2)$$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \ \ \bar{X} = 39.08, \ \ s_X^2 = 7.77$
  - Emergency: $m = 633, \ \ \bar{Y} = 39.60, \ \ s_Y^2 = 4.95$

$s_p^2 =$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

$$s_p^2 = \frac{774 \cdot 7.77 + 632 \cdot 4.95}{775 + 633 - 2}\left(\frac{1}{775} + \frac{1}{633}\right) = 0.0187$$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

$$T = \frac{\bar{X} - \bar{Y}}{s_p} =$$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

$$T = \frac{\bar{X} - \bar{Y}}{s_p} = \frac{39.08 - 39.60}{\sqrt{0.0187}} = -3.8064$$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

$$T = \frac{\bar{X} - \bar{Y}}{s_p} = \frac{39.08 - 39.60}{\sqrt{0.0187}} = -3.8064$$

$$T \sim t(1408) \approx N(0,1)$$

TWO-SAMPLE T-TEST: EXAMPLE

- Women admitted to maternity hospital, weeks of pregnancy measured in two groups.
  - Medical: $n = 775, \;\; \bar{X} = 39.08, \;\; s_X^2 = 7.77$
  - Emergency: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

$$H_0\colon \mu_x - \mu_Y = 0 - \text{rejected}$$

$$H_1\colon \mu_x - \mu_Y \neq 0$$

TWO-SAMPLE T-TEST: EXAMPLE

- Assumptions made:
  - Length of pregnancy is normally distributed
  - Variances are the same in both groups – *unrealistic*

TWO-SAMPLE T-TEST: EXAMPLE

- Assumptions made:
  - Length of pregnancy is normally distributed
  - Variances are the same in both groups – *unrealistic*

- **GOOD TO KNOW:**
  - There're special tests to check if the data is normal and if two groups have the same variances.
  - In practice, it's good to apply them first to check the validity of the assumptions.

WELCH'S TESTS

The case of unequal variances

WELCH'S TEST

- Two sets of data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$$

$$Y_1, Y_2, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$$

- Means μ_X and μ_Y and the variances σ_X^2 and σ_Y^2 are **unknown.**
  - Note that the variances aren't assumed to be equal.

WELCH'S TEST

- Two sets of data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \dots, X_n \sim N(\mu_X, \sigma_X^2)$$

$$Y_1, Y_2, \dots, Y_m \sim N(\mu_Y, \sigma_Y^2)$$

- Means μ_X and μ_Y and the variances σ_X^2 and σ_Y^2 are **unknown.**
  - Note that the variances aren't assumed to be equal.

- Null hypothesis: H_0: $\mu_X - \mu_Y = 0$

WELCH'S TEST

- Two sets of data drawn from normal distribution (i.i.d.):

$$X_1, X_2, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$$

$$Y_1, Y_2, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$$

- Means μ_X and μ_Y and the variances σ_X^2 and σ_Y^2 are **unknown.**
  - Note that the variances aren't assumed to be equal.
- Null hypothesis: $H_0\colon \mu_X - \mu_Y = 0$
- Alternatives:
  - Two-sided: $H_1\colon \mu_X - \mu_Y \neq 0$
  - One-sided: $H_1\colon \mu_X - \mu_Y > 0$ or $H_1\colon \mu_X - \mu_Y < 0$

WELCH'S TEST

- Test statistic:

$$T(X,Y) = \frac{\bar{X}-\bar{Y}}{s_p}\text{, where}$$

$$s_p^2 = \frac{s_X^2}{n} + \frac{s_Y^2}{m}, \qquad df = \frac{(s_X^2/n + s_Y^2/m)^2}{(s_X^2/n)^2/(n-1) + (s_Y^2/m)^2/(m-1)}$$

s_X^2, s_Y^2 - sample variances of X and Y.

WELCH'S TEST

- Test statistic:

$$T(X,Y) = \frac{\bar{X}-\bar{Y}}{s_p}\text{, where}$$

$$s_p^2 = \frac{s_X^2}{n} + \frac{s_Y^2}{m}, \qquad df = \frac{(s_X^2/n + s_Y^2/m)^2}{(s_X^2/n)^2/(n-1) + (s_Y^2/m)^2/(m-1)}$$

s_X^2, s_Y^2 - sample variances of X and Y.

- Null distribution:

$$\frac{\bar{X}-\bar{Y}}{s_p} \sim t(df)$$

PAIRED SAMPLE T-TEST

Difference between the means when data comes in pairs

PAIRED DATA

- Sometimes, data naturally comes in pairs $(X_i, Y_i), i = 1, \ldots, n$.

PAIRED DATA

- Sometimes, data naturally comes in pairs $(X_i, Y_i), i = 1, \ldots, n$.
- "Repeated measures":
  - patients' condition before and after taking a new medication;
  - students' performances before and after extra tutoring;
  - performance of two different algorithms on the same set of benchmarks.
- *Was there any effect?*

PAIRED DATA

- Sometimes, data naturally comes in pairs $(X_i, Y_i), i = 1, \dots, n$.
- "Repeated measures":
  - patients' condition before and after taking a new medication;
  - students' performances before and after extra tutoring;
  - performance of two different algorithms on the same set of benchmarks.
- *Was there any effect?*
- Samples are paired = dependent.

PAIRED SAMPLE T-TEST

- Two **<u>dependent</u>** sets of data drawn from normal distribution:

$$X_1, X_2, \dots, X_n, \qquad E(X_i) = \mu_X$$

$$Y_1, Y_2, \dots, Y_n, \qquad E(Y_i) = \mu_Y$$

PAIRED SAMPLE T-TEST

- Two **dependent** sets of data drawn from normal distribution:

$$X_1, X_2, \dots, X_n, \qquad E(X_i) = \mu_X$$

$$Y_1, Y_2, \dots, Y_n, \qquad E(Y_i) = \mu_Y$$

- Let's introduce $D = X_i - Y_i$ Assumption: $D \sim N(\mu, \sigma^2)$

PAIRED SAMPLE T-TEST

- Two **<u>dependent</u>** sets of data drawn from normal distribution:

$$X_1, X_2, \dots, X_n, \qquad E(X_i) = \mu_X$$

$$Y_1, Y_2, \dots, Y_n, \qquad E(Y_i) = \mu_Y$$

- Let's introduce $D = X_i - Y_i$ Assumption: $D \sim N(\mu, \sigma^2)$

FROM NOW, IT'S JUST A ONE-SAMPLE T-TEST IN TERMS OF D

PAIRED SAMPLE T-TEST

- Two <u>**dependent**</u> sets of data drawn from normal distribution:

$$X_1, X_2, \dots, X_n, \qquad E(X_i) = \mu_X$$

$$Y_1, Y_2, \dots, Y_n, \qquad E(Y_i) = \mu_Y$$

- Let's introduce $D = X_i - Y_i$ Assumption: $D \sim N(\mu, \sigma^2)$
- Null hypothesis:

$$H_0\colon E(D) = \mu = 0$$

- Alternatives:
  - Two-sided: $H_1: E(D) = \mu \neq 0$
  - One-sided: $H_1: \mu > 0$ or $H_1: \mu < 0$

PAIRED SAMPLE T-TEST

- Test statistic:

$$T(X,Y) = \frac{\overline{D}\sqrt{n}}{s},$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$$

PAIRED SAMPLE T-TEST

- Test statistic:

$$T(X,Y) = \frac{\bar{D}\sqrt{n}}{s},$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$$

- Null distribution:

$$T(X,Y) = \frac{\bar{D}\sqrt{n}}{s} \sim$$

PAIRED SAMPLE T-TEST

- Test statistic:

$$T(X,Y) = \frac{\bar{D}\sqrt{n}}{s},$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$$

- Null distribution:

$$T(X,Y) = \frac{\bar{D}\sqrt{n}}{s} \sim t(n-1)$$