PROBABILITY AND STATISTICS

Lecture 13 – Hypothesis testing

HYPOTHESIS TESTING

HYPOTHESIS TESTING

- After tossing a coin 100 times, we observed H only 30 times. Is it a fair coin?
- Did extra tuition help students?
- Is drug A more efficient than drug B?

HOW IT ALL STARTED

https://youtu.be/lgs7d5saFFc

LADY TASTING TEA

Ronald Fisher, 1913

LADY TASTING TEA

- 8 cups of tea
  - 4 cups: milk first
  - 4 cups: tea first

- The lady must select 4 cups prepared by one method.

- *How to check her ability to distinguish the teas?*

Ronald Fisher, 1913

LADY TASTING TEA

- The default assumption:

H_0: the lady can't distinguish the teas

LADY TASTING TEA

- The default assumption:

H_0: the lady can't distinguish the teas

- How unlikely would it be to randomly guess all 4?

LADY TASTING TEA

- The default assumption:

H_0: the lady can't distinguish the teas

- How unlikely would it be to randomly guess all 4?

Tea-Tasting Distribution Assuming H_0		
Success count	Combinations of selection	Number of Combinations
0	oooo	1 × 1 = 1
1	ooox, ooxo, oxoo, xooo	4 × 4 = 16
2	ooxx, oxox, oxxo, xoxo, xxoo, xoox	6 × 6 = 36
3	oxxx, xoxx, xxox, xxxo	4 × 4 = 16
4	xxxx	1 × 1 = 1
Total		70

LADY TASTING TEA

- The default assumption:

H_0: the lady can't distinguish the teas

- How unlikely would it be to randomly guess all 4?

$$1/70 \approx 0.014$$

Tea-Tasting Distribution Assuming H_0		
Success count	Combinations of selection	Number of Combinations
0	oooo	1 × 1 = 1
1	ooox, ooxo, oxoo, xooo	4 × 4 = 16
2	ooxx, oxox, oxxo, xoxo, xxoo, xoox	6 × 6 = 36
3	oxxx, xoxx, xxox, xxxo	4 × 4 = 16
4	xxxx	1 × 1 = 1
Total		70

LADY TASTING TEA

- The default assumption:

H_0: the lady can't distinguish the teas

- How unlikely would it be to randomly guess all 4?

$$1/70 \approx 0.014$$

That's *surprising enough* to reject H_0

Tea-Tasting Distribution Assuming H_0		
Success count	Combinations of selection	Number of Combinations
0	oooo	$1 \times 1 = 1$
1	ooox, ooxo, oxoo, xooo	$4 \times 4 = 16$
2	ooxx, oxox, oxxo, xoxo, xxoo, xoox	$6 \times 6 = 36$
3	oxxx, xoxx, xxox, xxxo	$4 \times 4 = 16$
4	xxxx	$1 \times 1 = 1$
Total		70

HYPOTHESIS TESTING

- Test whether a default assumption about the data is plausible.

- Key idea: you can't *prove* something, but you can disprove it.

HYPOTHESIS TESTING

- Test whether a default assumption about the data is plausible.

HYPOTHESIS TESTING

ASSUMPTION:
ALL THE SWANS ARE WHITE

HYPOTHESIS TESTING

ASSUMPTION:
ALL THE SWANS ARE WHITE

EVIDENCE:

HYPOTHESIS TESTING

ASSUMPTION:
ALL THE SWANS ARE WHITE

EVIDENCE:
NOT CONVINCING…

HYPOTHESIS TESTING

ASSUMPTION:
ALL THE SWANS ARE WHITE

EVIDENCE:
NOW IT's CONVINCING!

HYPOTHESIS TESTING

ASSUMPTION:
ALL THE SWANS ARE WHITE

EVIDENCE:

HYPOTHESIS TESTING

ASSUMPTION:
THERE'RE NO ALIENS ON THE MOON

HYPOTHESIS TESTING

ASSUMPTION:
THERE'RE NO ALIENS ON THE MOON

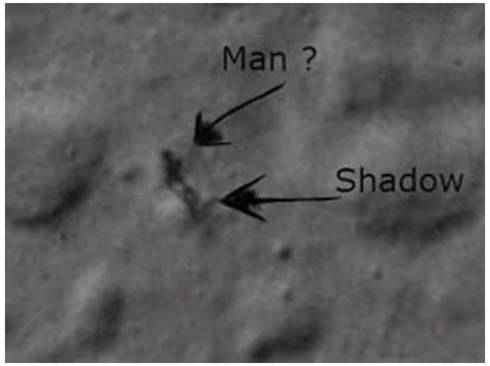
EVIDENCE:

HYPOTHESIS TESTING

ASSUMPTION:
THERE'RE NO ALIENS ON THE MOON

EVIDENCE:

STILL NO REASON TO BELIEVE THERE ARE ALIENS

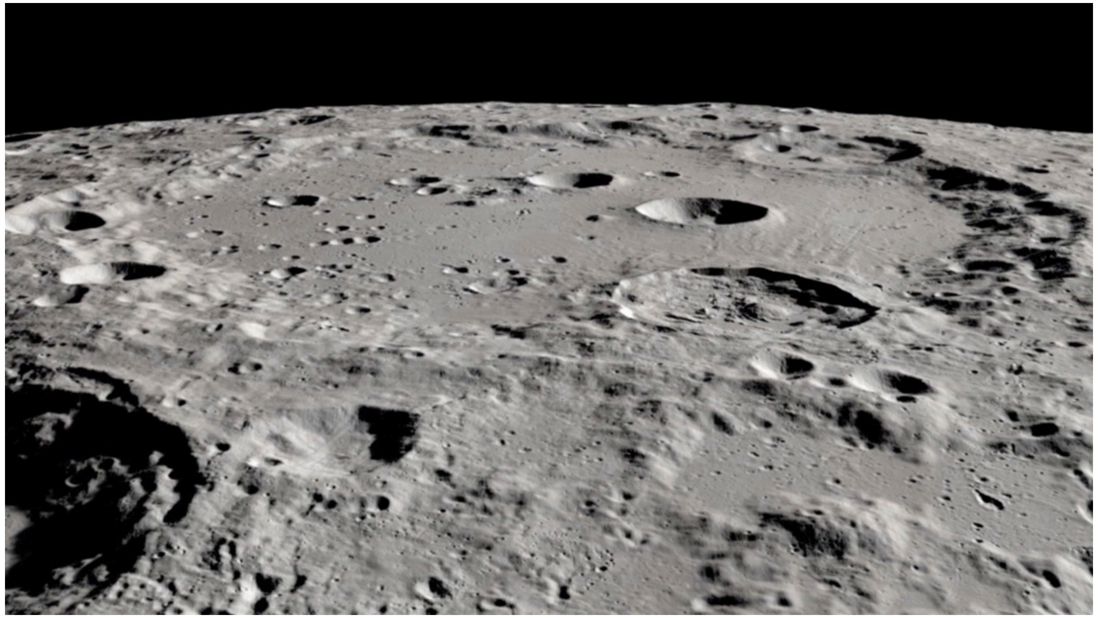


HYPOTHESIS TESTING

ASSUMPTION:
THERE'RE NO ALIENS ON THE MOON

EVIDENCE:

STILL NO REASON TO BELIEVE THERE ARE ALIENS



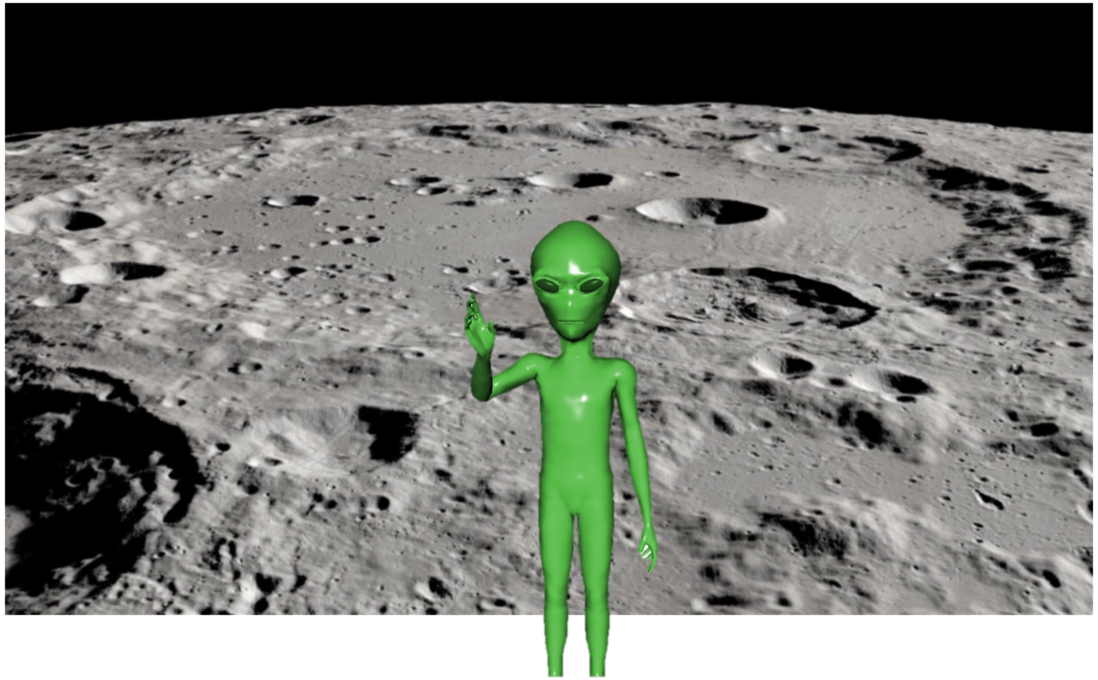
HYPOTHESIS TESTING

ASSUMPTION:

THERE'RE NO ALIENS ON THE MOON

EVIDENCE:

THERE ARE ALIENS!!!



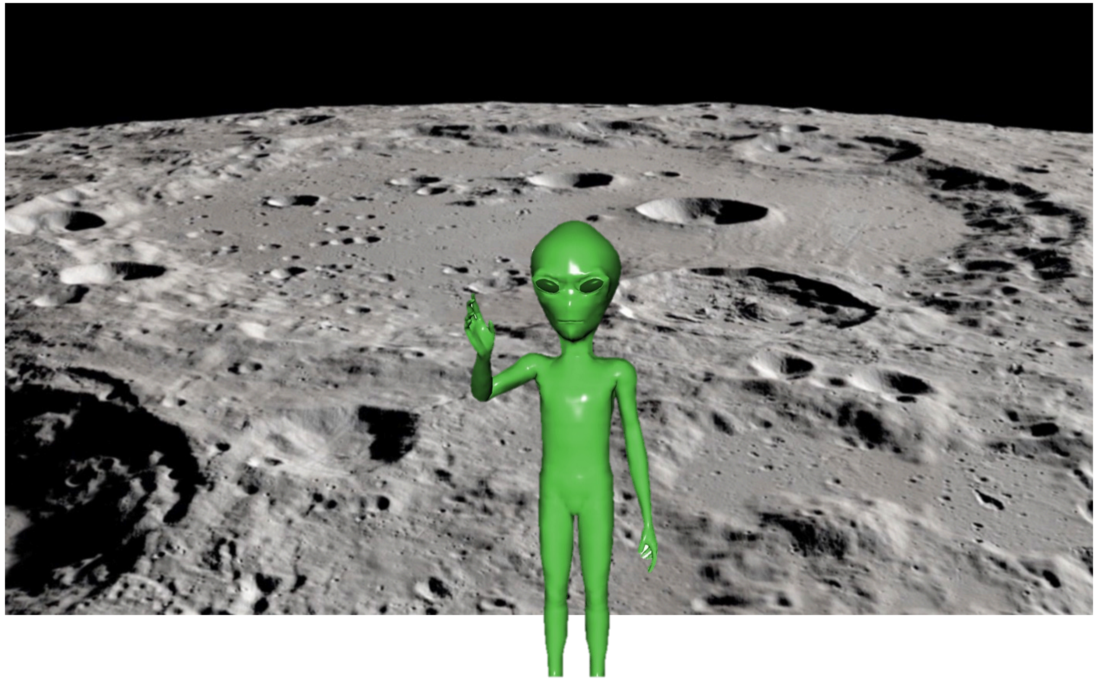
HYPOTHESIS TESTING

ASSUMPTION:

THERE'RE NO ALIENS ON THE MOON

EVIDENCE:

THERE ARE ALIENS!!!



HYPOTHESIS TESTING

- Note:
  - Default assumption = “status quo”.

HYPOTHESIS TESTING

- Note:
  - Default assumption = “status quo”.
  - Alternative: something we can find evidence for.

HOW TO TEST A HYPOTHESIS

BASIC INGREDIENTS

RUNNING EXAMPLE

How to check if a coin is a fair one?

RUNNING EXAMPLE

We flip a coin 10 times to test if it's a fair one.

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis** H_0 : *the default assumption for the model generating the data.*

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis $\boldsymbol{H_0}$** : *the default assumption for the model generating the data.*

$$H_0\colon p = 0.5 \text{ ‘the coin is fair’}$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis $\boldsymbol{H_0}$** : *the default assumption for the model generating the data.*

$$H_0\colon p = 0.5 \text{ 'the coin is fair'}$$

- **The alternative hypothesis $\boldsymbol{H_1}$**: *if we reject H_0, we accept H_1 as the best explanation for the data.*

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis $\boldsymbol{H_0}$** : *the default assumption for the model generating the data.*

$$H_0 \colon p = 0.5 \ \text{'the coin is fair'}$$

- **The alternative hypothesis $\boldsymbol{H_1}$**: *if we reject H_0, we accept H_1 as the best explanation for the data.*

$$H_1 \colon p \neq 0.5 \ \text{'the coin is not fair'}$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis $\boldsymbol{H_0}$** : *the default assumption for the model generating the data.*

$$H_0\colon\ p = 0.5\ \text{'the coin is fair'}$$

- **The alternative hypothesis $\boldsymbol{H_1}$**: *if we reject H_0, we accept H_1 as the best explanation for the data.*

$$H_1\colon\ p \neq 0.5\ \text{'the coin is not fair'}$$

$$H_1\colon p > 0.5$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis $\boldsymbol{H_0}$** : *the default assumption for the model generating the data.*

$$H_0\colon\ p = 0.5 \ \text{‘the coin is fair’}$$

- **The alternative hypothesis $\boldsymbol{H_1}$**: *if we reject H_0, we accept H_1 as the best explanation for the data.*

$$H_1\colon\ p \neq 0.5 \ \text{‘the coin is not fair’}$$

$$H_1\colon p > 0.5 \quad \text{or} \quad H_1\colon p < 0.5$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis $\boldsymbol{H_0}$** : *the default assumption for the model generating the data.*

$$H_0 \colon p = 0.5 \text{ ‘the coin is fair’}$$

- **The alternative hypothesis $\boldsymbol{H_1}$**: *if we reject H_0, we accept H_1 as the best explanation for the data.*

Two-sided: $H_1 \colon p \neq 0.5$ ‘the coin is not fair’

One-sided: $H_1 \colon p > 0.5$ or $H_1 \colon p < 0.5$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

- **The null hypothesis $\boldsymbol{H_0}$** : *the default assumption for the model generating the data.*

$$H_0 : p = 0.5 \text{ 'the coin is fair'}$$

- **The alternative hypothesis $\boldsymbol{H_1}$**: *if we reject H_0, we accept H_1 as the best explanation for the data.*

Two-sided: $H_1 : p \neq 0.5$ 'the coin is not fair'

One-sided: $H_1 : p > 0.5$ or $H_1 : p < 0.5$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \textit{number of H}$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \text{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \text{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \textit{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim Binomial(\qquad)$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \text{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim Binomial(10, 0.5)$$

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \textit{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim Binomial(10, 0.5)$$

X											
$P(X\|H_0)$											

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \textit{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim Binomial(10, 0.5)$$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X \vert H_0)$											

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \textit{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim Binomial(10, 0.5)$$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X \vert H_0)$	0.001										0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - number\ of\ H$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim Binomial(10, 0.5)$$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01								0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

- **Test statistic**: *computed from the data*

$$X - \textit{number of H}$$

- **Null distribution**: *the probability distribution of* X *assuming* H_0

$$X \sim Binomial(10, 0.5)$$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

- **Rejection region R**: if X is in the rejection region, we reject H_0 in favor of H_1.

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

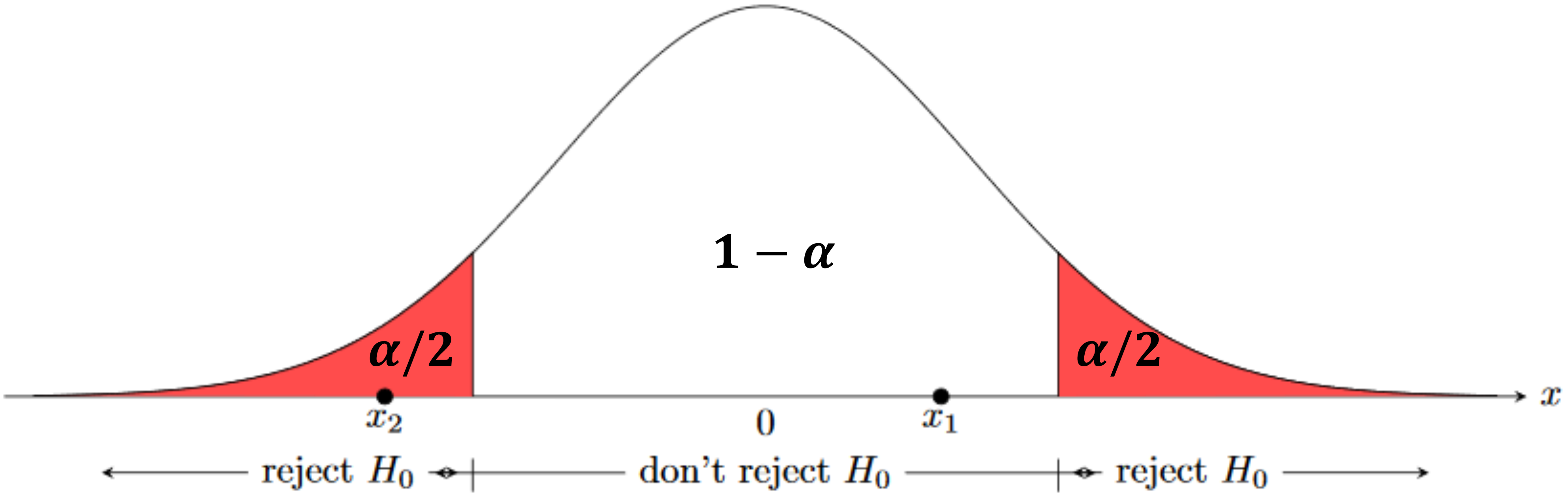
H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

- **Rejection region R**: if X is in the rejection region, we reject H_0 in favor of H_1.
- **Significance level** α **:** $P(X \in R|H_0) \leq \alpha$
  - Typically chosen in advance (common values are 0.1, 0.05, 0.01.)

INGREDIENTS

- Distribution of the test statistic under H_0, two-sided alternative:



INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

$$\boldsymbol{\alpha = 0.05}$$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

$$\boldsymbol{\alpha = 0.05}$$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X \mid H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

$$\boldsymbol{\alpha = 0.2}$$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

$$\boldsymbol{\alpha = 0.2}$$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

REMARKS

- **The null hypothesis** is the *cautious default*
  - we won’t claim the coin is unfair unless we have compelling evidence.

REMARKS

- **The null hypothesis** is the *cautious default*
  - we won't claim the coin is unfair unless we have compelling evidence.

- **Rejection region**: data that is *extreme* under the null hypothesis = outcomes in the tail(s) of the null distribution.
  - depends on the significance level α of the test.

REMARKS

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair' $\boldsymbol{\alpha = 0.05}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

REMARKS

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair' $\boldsymbol{\alpha = 0.05}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

- If we get *0, 1, 9, 10* H
  - Reject H_0

REMARKS

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair' $\boldsymbol{\alpha = 0.05}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

- If we get *0, 1, 9, 10* H
  - Reject H_0

- If we get 2, *3, 4, 5, 6, 7, 8* H, then the test statistic is in the non-rejection region.

REMARKS

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair' $\boldsymbol{\alpha = 0.05}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

- If we get *0, 1, 9, 10* H
  - Reject H_0
- If we get 2, *3, 4, 5, 6, 7, 8* H, then the test statistic is in the non-rejection region.
  - Interpretation: the data 'does not support rejecting the null hypothesis'.

REMARKS

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair' $\boldsymbol{\alpha = 0.05}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

- If we get *0, 1, 9, 10* H
  - Reject H_0
- If we get 2, *3, 4, 5, 6, 7, 8* H, then the test statistic is in the non-rejection region.
  - Interpretation: the data 'does not support rejecting the null hypothesis'.
  - **Never claim that the data proves the null hypothesis is true.**

EXPERIMENT

EXPERIMENT

- Toss a coin 10 times, record the number of heads you got.

EXPERIMENT

- Toss a coin 10 times, record the number of heads you got.

- Based on the result, do you reject the null hypothesis?

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair', $\alpha = 0.2$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INCORRECTLY REJECTING H_0

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

What is the probability to incorrectly reject H_0?

INCORRECTLY REJECTING H_0

H_0: $p = 0.5$ 'the coin is fair', H_1: $p \neq 0.5$ 'the coin is not fair'

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

What is the probability to incorrectly reject H_0?

$$P(Reject\ H_0|H_0) = P(X \in R|H_0) = \alpha -$$

significance level

ONE-SIDED ALTERNATIVES

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$, $\boldsymbol{H_1}$: $\boldsymbol{p > 0.5}$ 'the biased towards H'

$$\boldsymbol{\alpha = 0.05}$$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$, $\boldsymbol{H_1}$: $\boldsymbol{p > 0.5}$ 'the biased towards H'

$\boldsymbol{\alpha = 0.05}$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$, $\boldsymbol{H_1}$: $\boldsymbol{p > 0.5}$ 'the biased towards H'

$\boldsymbol{\alpha = 0.1}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$, $\boldsymbol{H_1}$: $\boldsymbol{p > 0.5}$ 'the biased towards H'

$\boldsymbol{\alpha = 0.1}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$, $\boldsymbol{H_1}$: $\boldsymbol{p < 0.5}$ 'the biased towards T'

$\boldsymbol{\alpha = 0.05}$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

$H_0\colon p = 0.5,$ $\boldsymbol{H_1\colon p < 0.5}$ 'the biased towards T'

$\boldsymbol{\alpha = 0.05}$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$, $\boldsymbol{H_1}$: $\boldsymbol{p < 0.5}$ 'the biased towards T'

$\boldsymbol{\alpha = 0.1}$

X	0	1	2	3	4	5	6	7	8	9	10
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

INGREDIENTS

We flip a coin 10 times to test if it's a fair one.

H_0: $p = 0.5$, $\boldsymbol{H_1}$: $\boldsymbol{p < 0.5}$ 'the biased towards T'

$$\boldsymbol{\alpha = 0.1}$$

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P(X\|H_0)$	0.001	0.01	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.01	0.001

Z-TEST

When the data is coming from normal distribution
and variance is known

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.
- We suspect that HarbourSpace students are different. How do we test it?

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.

- We suspect that HarbourSpace students are different. How do we test it?

$$H_0:\ X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} = \mu_0$$

'HarbourSpace students' IQs are the same as those of the general population'

$$H_1: X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} > \mu_0$$

'HarbourSpace students' IQs are higher than those of the general population'

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.

- We suspect that HarbourSpace students are different. How do we test it?

$$H_0: \ X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} = \mu_0$$

'HarbourSpace students' IQs are the same as those of the general population'

$$H_1: X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} \neq \mu_0$$

'HarbourSpace students' IQs aren't the same as than those of the general population'

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.

- We suspect that HarbourSpace students are different. How do we test it?

$$H_0:\ X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} = \mu_0$$

'HarbourSpace students' IQs are the same as those of the general population'

$$H_1: X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} < \mu_0$$

'HarbourSpace students' IQs are lower than those of the general population'

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.
- We suspect that HarbourSpace students are different. How do we test it?

$$H_0\colon\ X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} = \mu_0$$

'HarbourSpace students' IQs are the same as those of the general population'

$$H_1\colon X \sim N(\mu_{HS}, \sigma^2), \qquad \mu_{HS} \neq \mu_0$$

'HarbourSpace students' IQs aren't the same as than those of the general population'

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.
- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\qquad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.
- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Average IQ of n students: $\bar{X}$

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.
- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\quad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Average IQ of n students: $\bar{X}$
- Significance level α

Z-TEST

- IQ follows $N(\mu_0, \sigma^2)$ where σ is known.
- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\qquad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Average IQ of n students: $\bar{X}$
- Significance level α
- Can we reject H_0?

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\quad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\qquad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\qquad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Assuming H_0, $X_i \sim$

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\qquad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Assuming $H_0,\ \ X_i \sim N(\mu_0, \sigma^2)$

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\qquad$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Assuming H_0, $X_i \sim N(\mu_0, \sigma^2)$

$$\Rightarrow \qquad \qquad \sim N(0,1)$$

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0 \qquad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Assuming $H_0,\ \ X_i \sim N(\mu_0, \sigma^2)$

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

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0 \qquad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Assuming H_0, $X_i \sim N(\mu_0, \sigma^2)$

$$\Rightarrow T(X_1, \ldots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$$

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0 \qquad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Assuming $H_0,\ \ X_i \sim N(\mu_0, \sigma^2)$

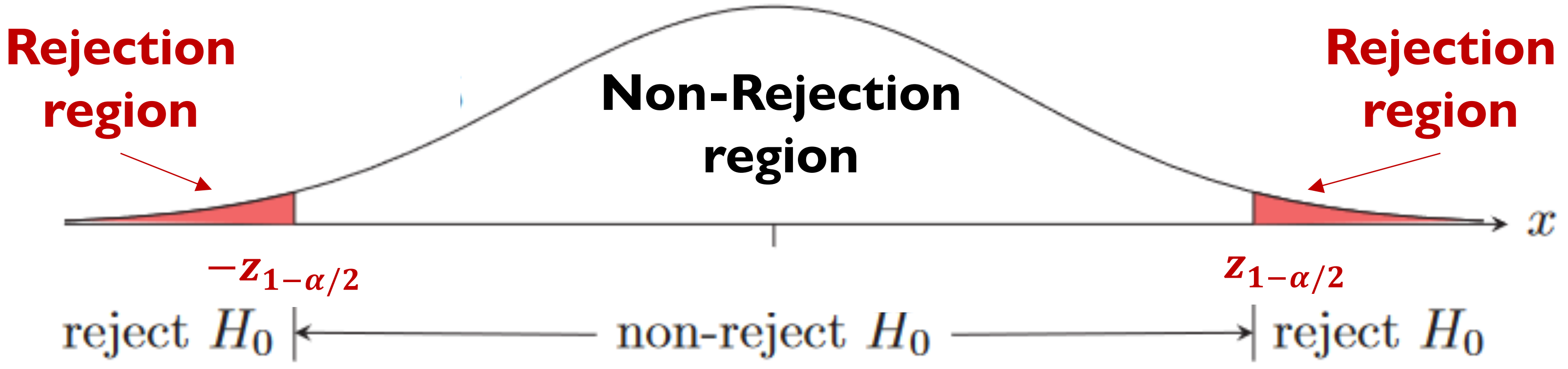
$$\Longrightarrow T(X_1, \dots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$$

Z-TEST

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0 \quad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic: $T(X_1, \dots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$
- Significance level α
- (non-) Rejection region:

Z-TEST

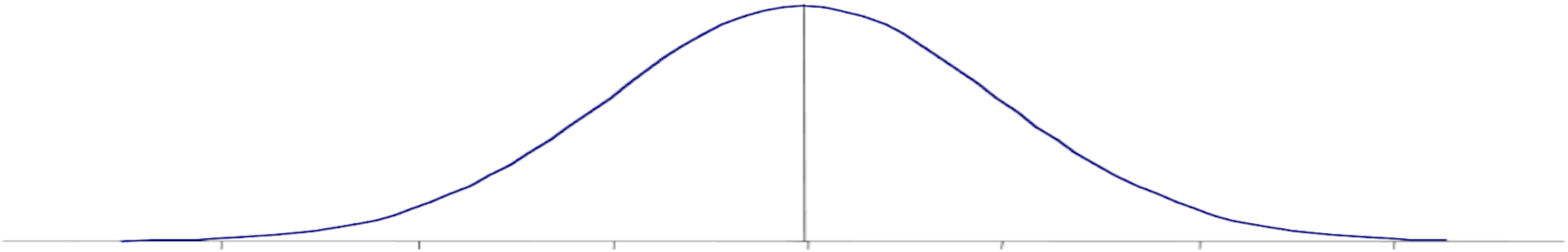
- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0 \quad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu \neq \mu_0$
- Test statistic: $T(X_1, \ldots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$
- Significance level α
- (non-) Rejection region:



ONE-SIDED ALTERNATIVES

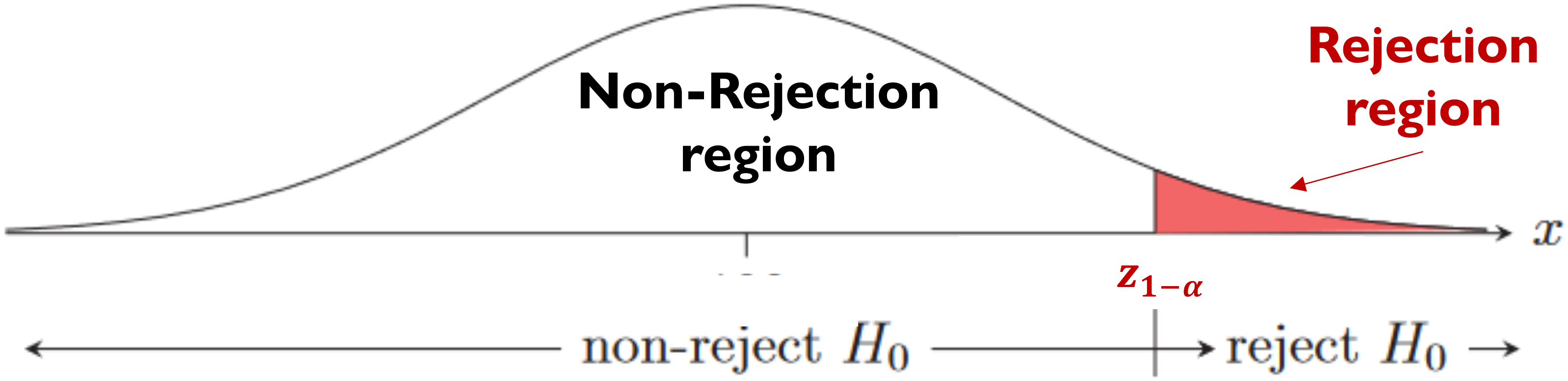
Z-TEST: ONE-SIDED

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\quad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu > \mu_0$
- Test statistic: $T(X_1, \ldots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$
- Significance level α
- (non-) Rejection region:



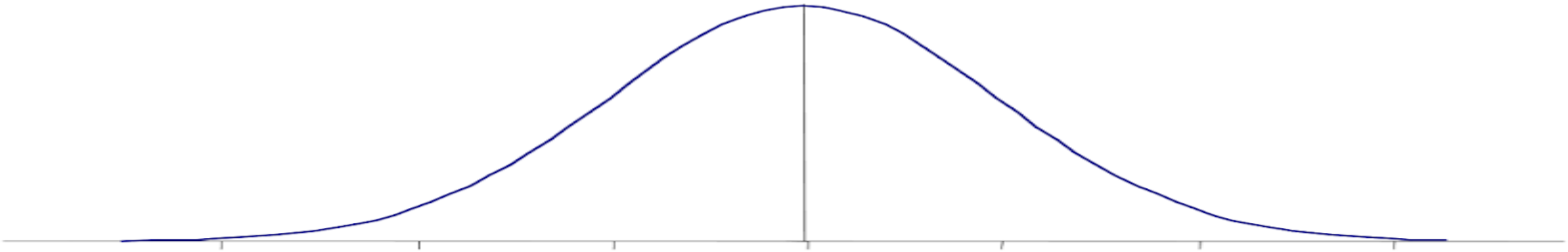
Z-TEST: ONE-SIDED

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\quad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu > \mu_0$
- Test statistic: $T(X_1, \dots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$
- Significance level α
- (non-) Rejection region:



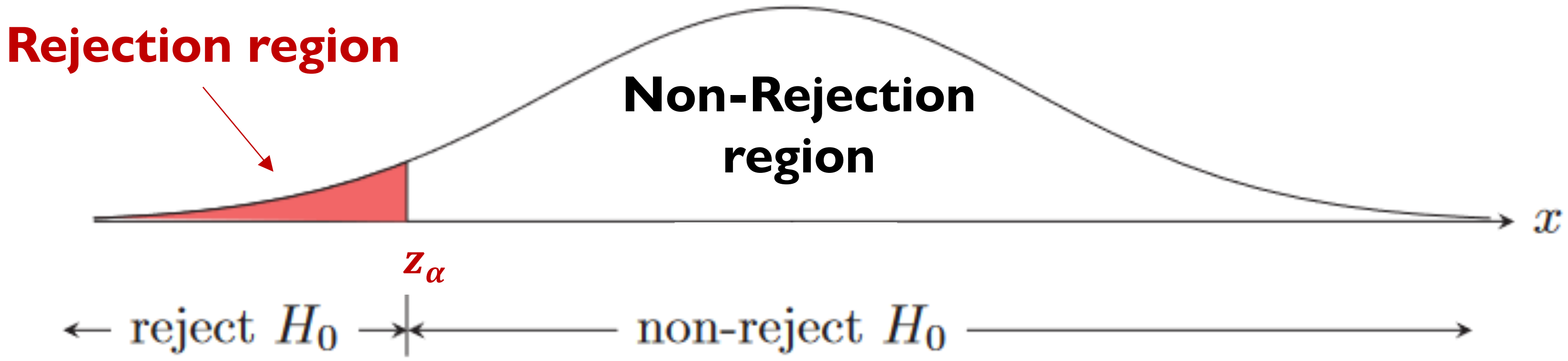
Z-TEST: ONE-SIDED

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu < \mu_0$
- Test statistic: $T(X_1, \ldots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$
- Significance level α
- (non-) Rejection region:



Z-TEST: ONE-SIDED

- $H_0: X \sim N(\mu_{HS}, \sigma^2),\ \mu = \mu_0$ $\quad H_1: X \sim N(\mu_{HS}, \sigma^2),\ \mu < \mu_0$
- Test statistic: $T(X_1, \ldots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1)$
- Significance level α
- (non-) Rejection region:



Z-TEST: EXAMPLE 1

- Weight loss is described by $X \sim N(5, 7^2)$.
- $n = 30$ patients followed an experimental diet.
- Average weight loss $\bar{X} = 6.1$ kg.

Does the diet make any difference?

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 $\alpha = 0.05$.

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

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

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

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 $\alpha = 0.05$.

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

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

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

Z-TEST: EXAMPLE 2

- A bakery supplies loaves of bread to supermarkets. Not every loaf weighs the same: loaf weight $X \sim N(\mu_0, 0.1^2)$, supermarket expects $\mu_0 = 2$ kg.
- A supermarket draws a sample of $n = 20$ loaves: average weights is $\bar{X} = 1.97$ kg.
- The supermarket wants to be sure that the weights are, on average, not lower than 2 kg.

Is there evidence against this?

T-TEST

When the data is coming from normal distribution
and variance is unknown

T-TEST

- Hypothesis about μ from $X \sim N(\mu_0, \sigma^2)$, σ is unknown.

T-TEST

- Hypothesis about μ from $X \sim N(\mu_0, \sigma^2)$, σ is unknown.

- Unknown variance is estimated from the sample:

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

T-TEST

- Hypothesis about μ from $X \sim N(\mu_0, \sigma^2)$, σ is unknown.

- Unknown variance is estimated from the sample:

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

- The test statistic is therefore:

$$T(X_1, \dots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim$$

T-TEST

- Hypothesis about μ from $X \sim N(\mu_0, \sigma^2)$, σ is unknown.

- Unknown variance is estimated from the sample:

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

- The test statistic is therefore:

$$T(X_1, \dots, X_n) = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \sim t(n-1)$$

T-TEST: EXAMPLE 2

- Sales at your company: $X \sim N(\mu_0, \sigma^2)$, $\mu_0 = 100$ dollars per transaction.

- You hire new employees. 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 new employees better than the old ones?
Test your hypothesis at $\alpha = 0.05$.

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 $N(\mu, \sigma^2)$.

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

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

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

TYPES OF ERRORS

TYPES OF ERRORS

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

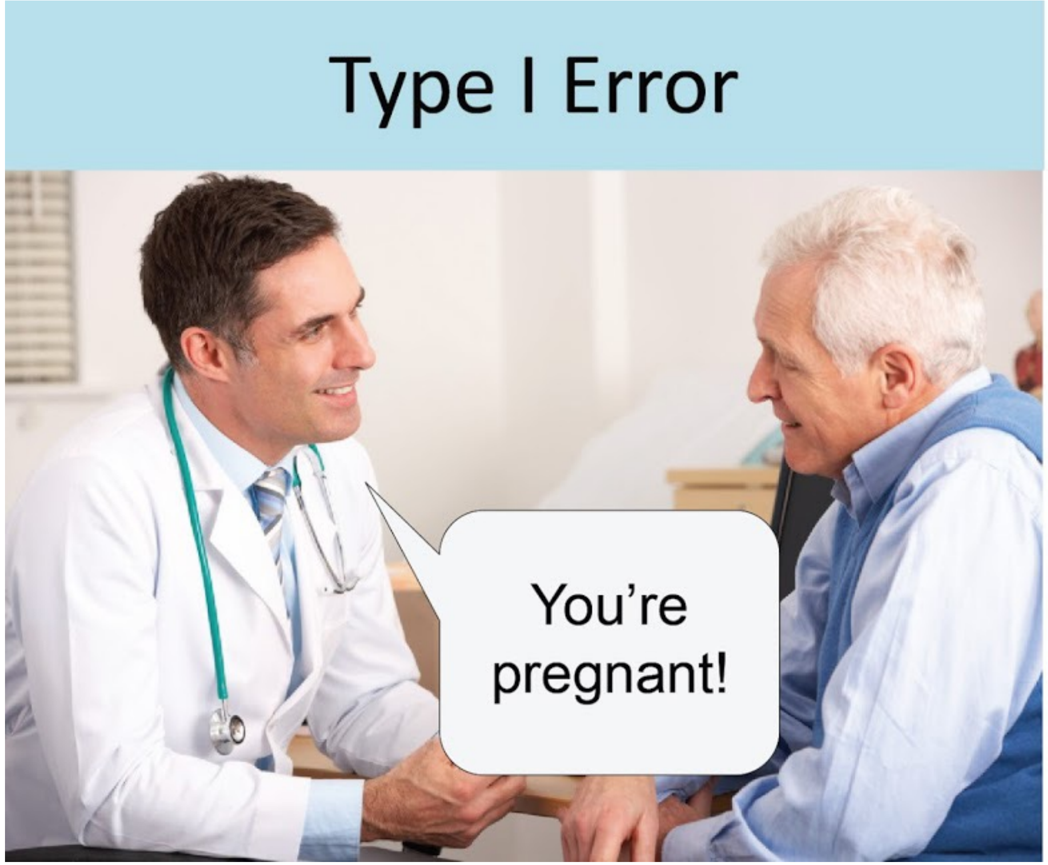
TYPES OF ERRORS

		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



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

P-VALUES

P-VALUES

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

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

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

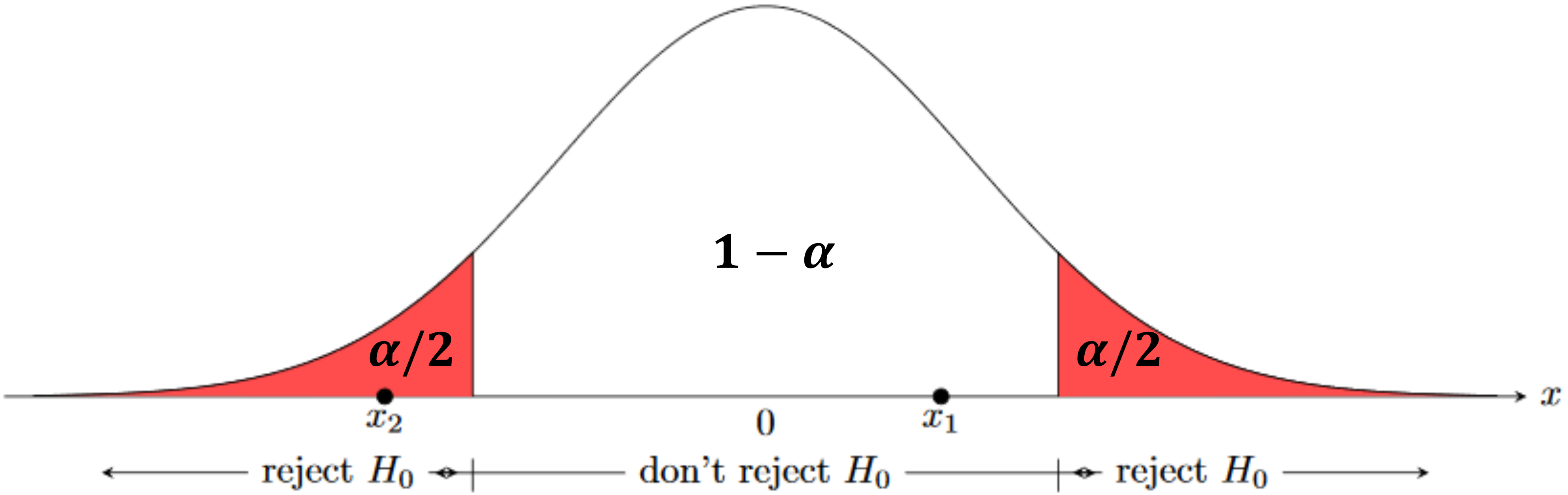
- Up till now:
  - Rejecting H_0 based on the value of the test statistic:

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

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

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

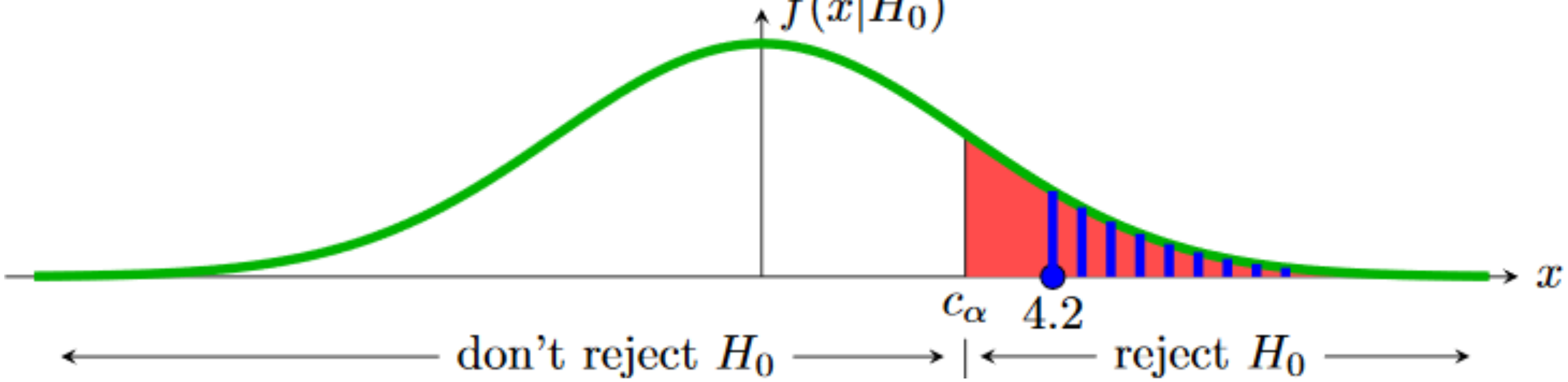
P-VALUES



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

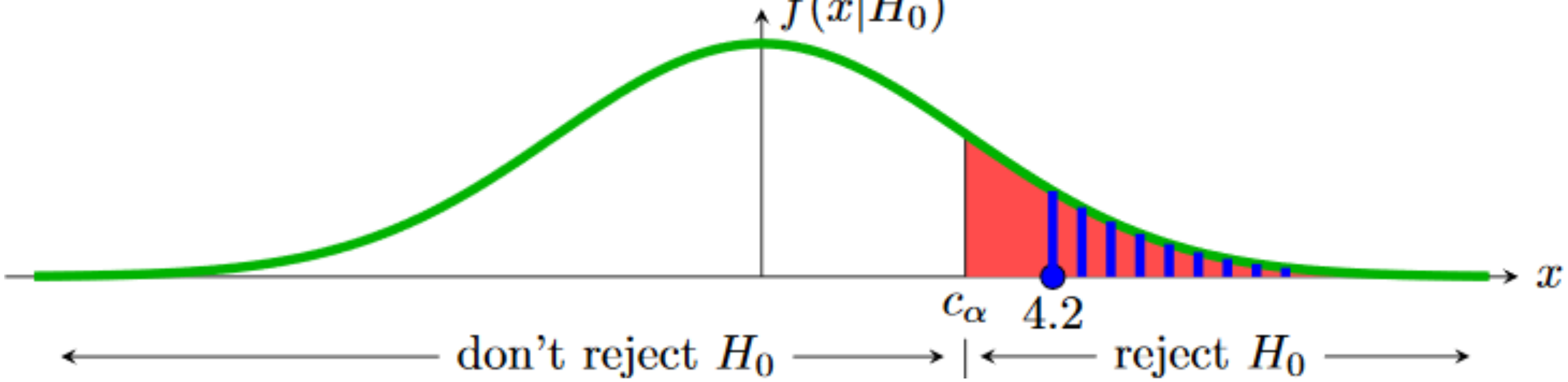
P-VALUES

- 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

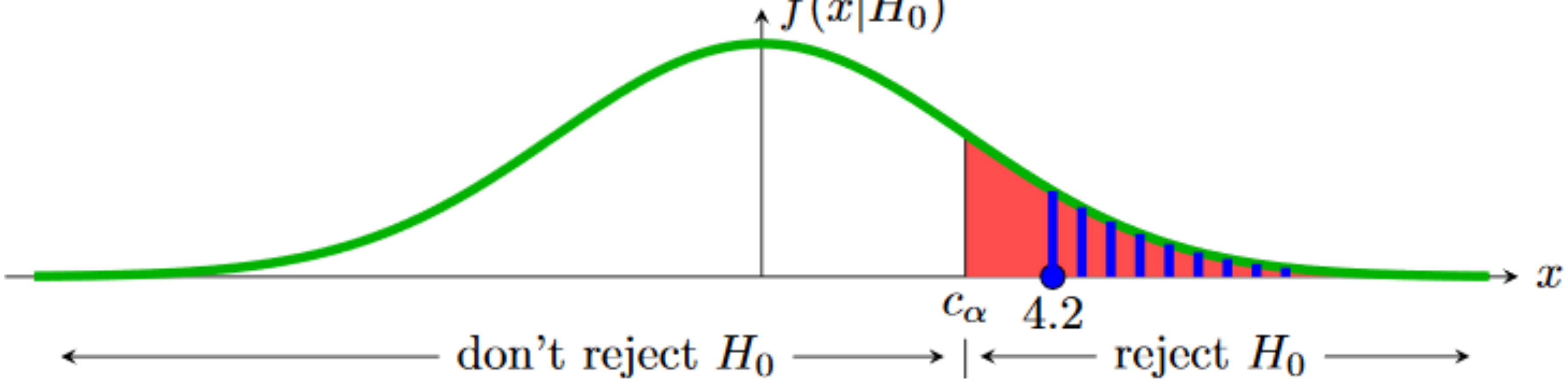
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t = 4.2. Should we reject H_0?



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

P-VALUES

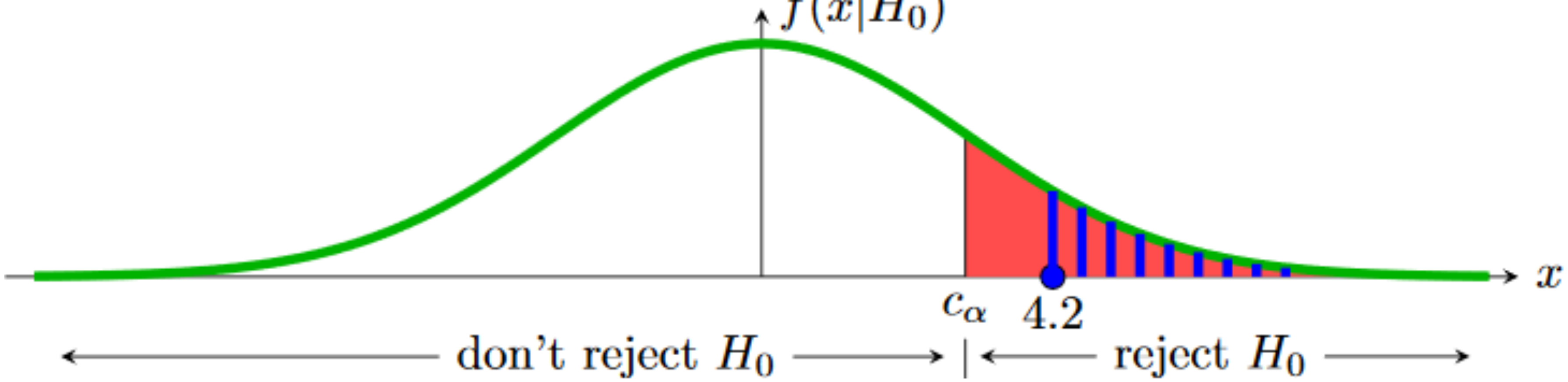
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t = 4.2. Should we reject H_0?



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

P-VALUES

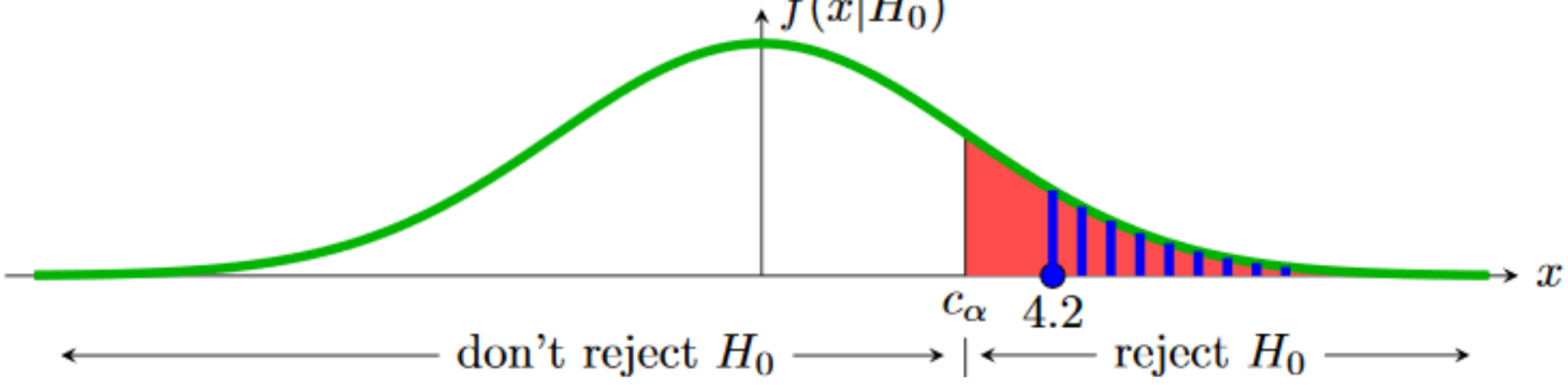
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic $t = 4.2$. Should we reject H_0?



- 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?

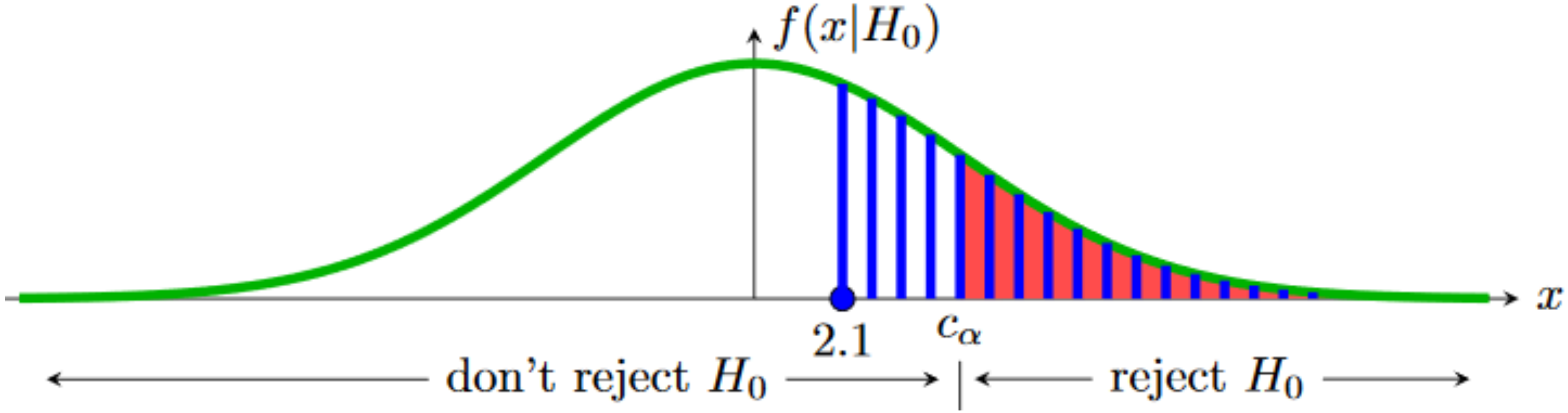


- Significance: $\alpha = P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area

Since $p < \alpha$, reject H_0

P-VALUES

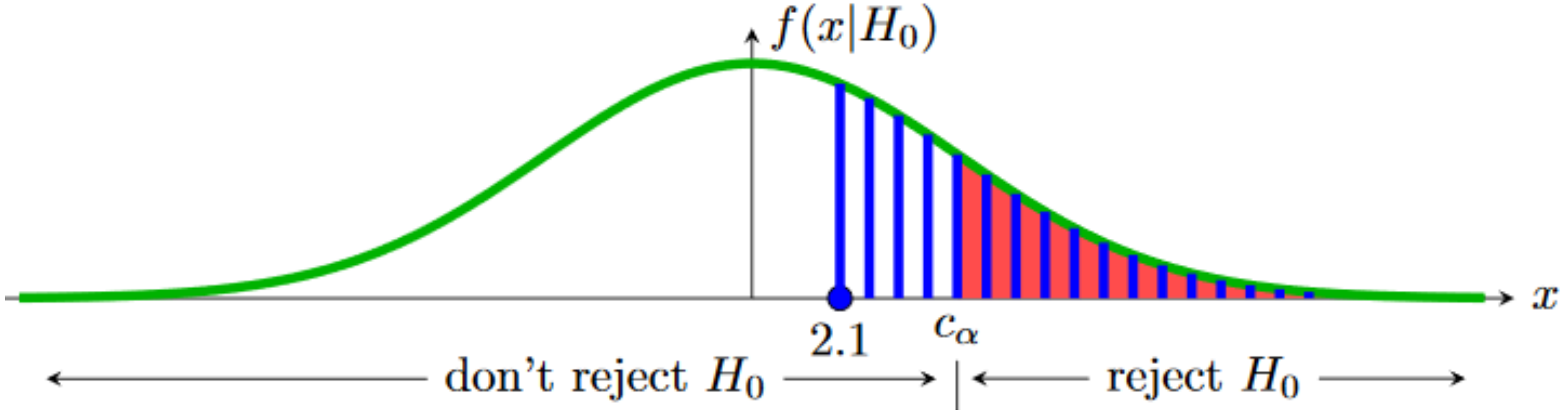
- Example:
  - Suppose we have the right-sided rejection region. We see data with test statistic t $= 2.1$. Should we reject H_0?



- 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 $\mathrm{t} = 2.1$. Should we reject H_0?

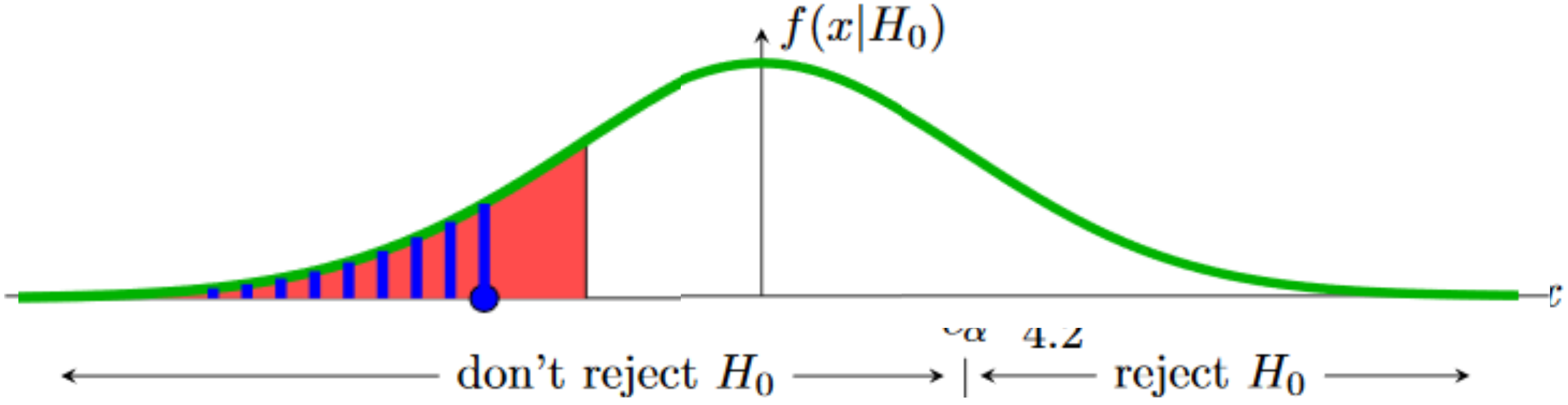


- Significance: $\alpha = P(T \in R|H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T}|H_0)$ = blue area

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

P-VALUES

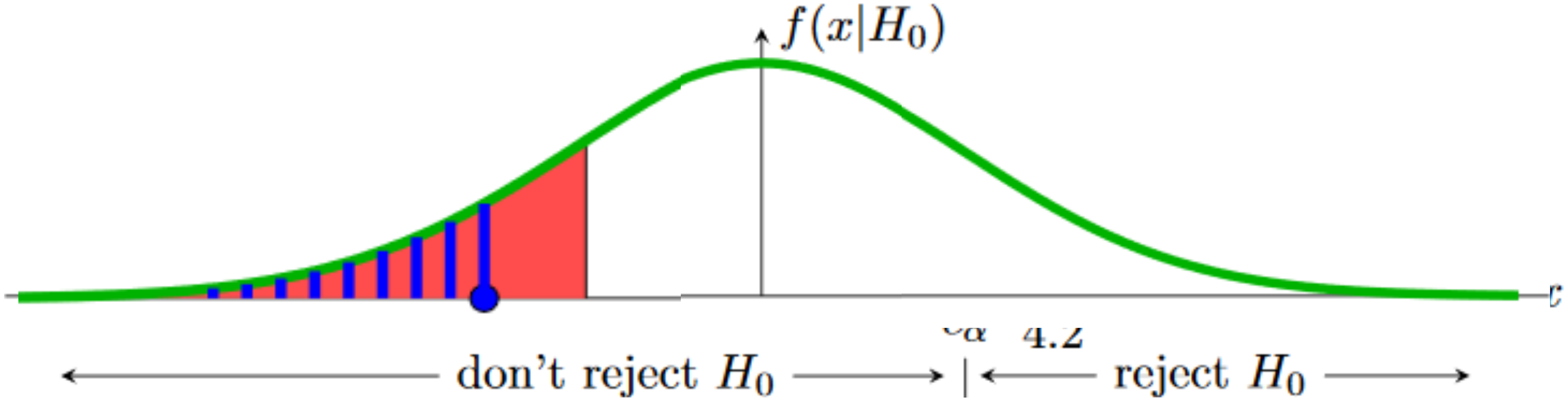
- Similarly, with the other one-sided alternative:
- Significance: $\alpha = P(T \in R | H_0)$ = red area.
- p-value: $p = P(\text{data at least as extreme as T} | H_0)$ = blue area



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

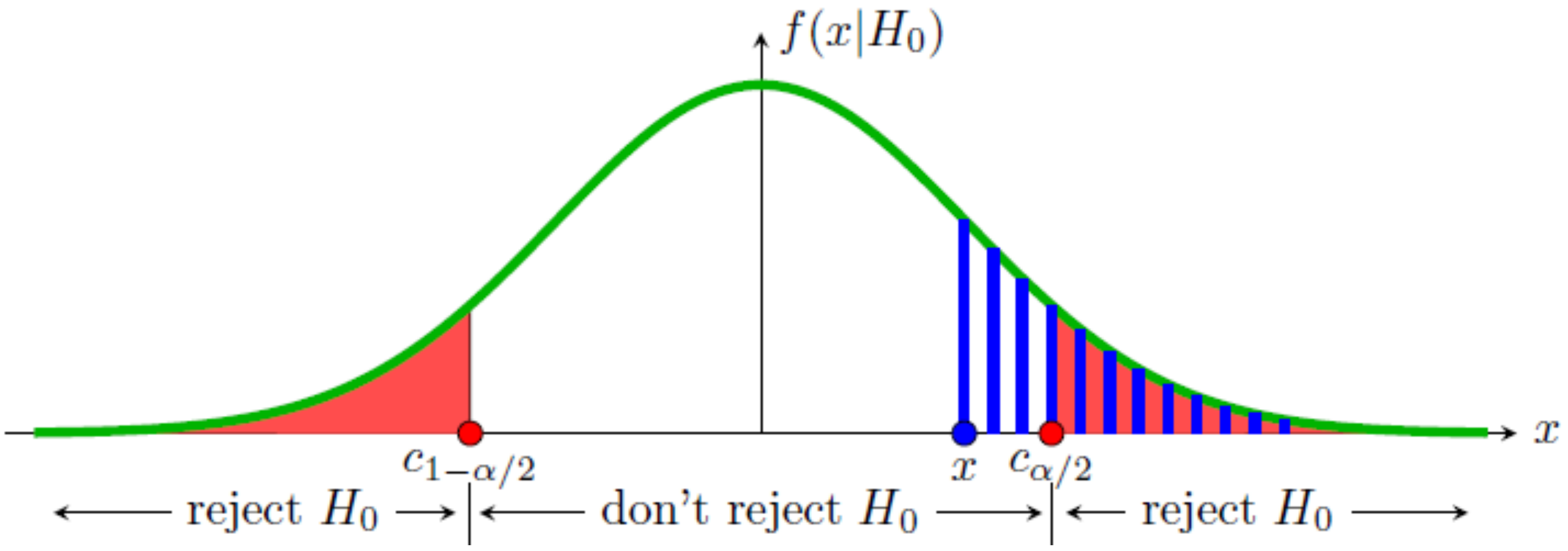


P-VALUES

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

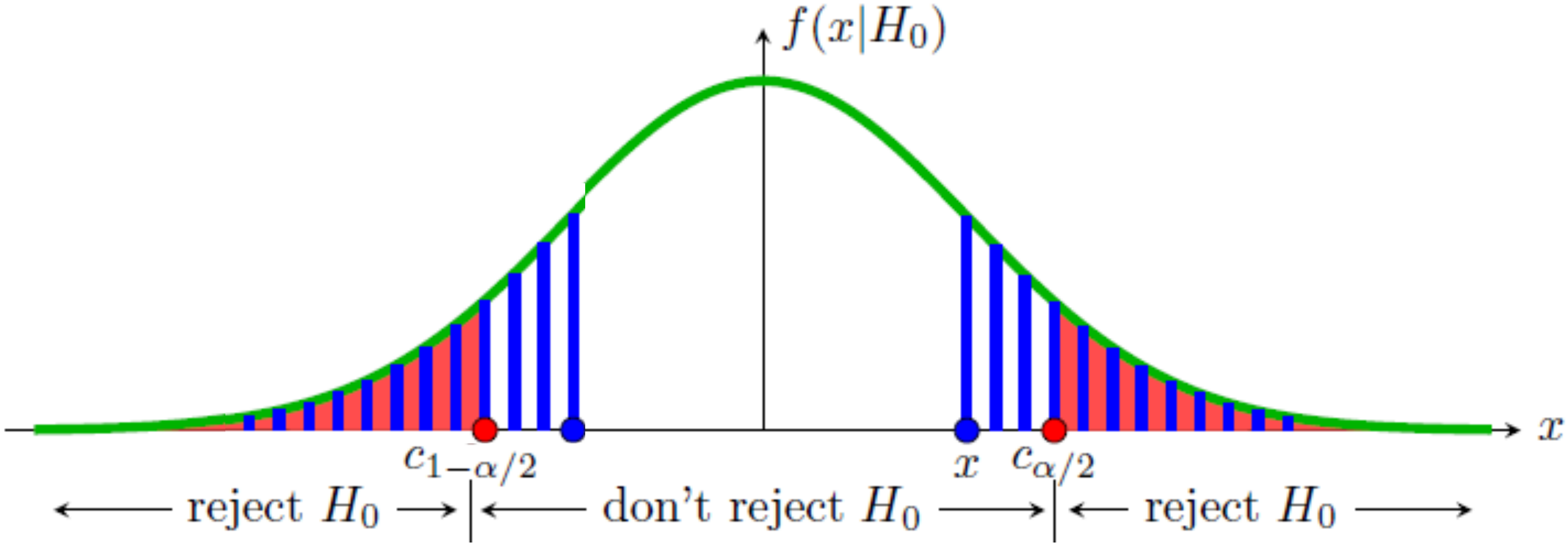
P-VALUES

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



P-VALUES

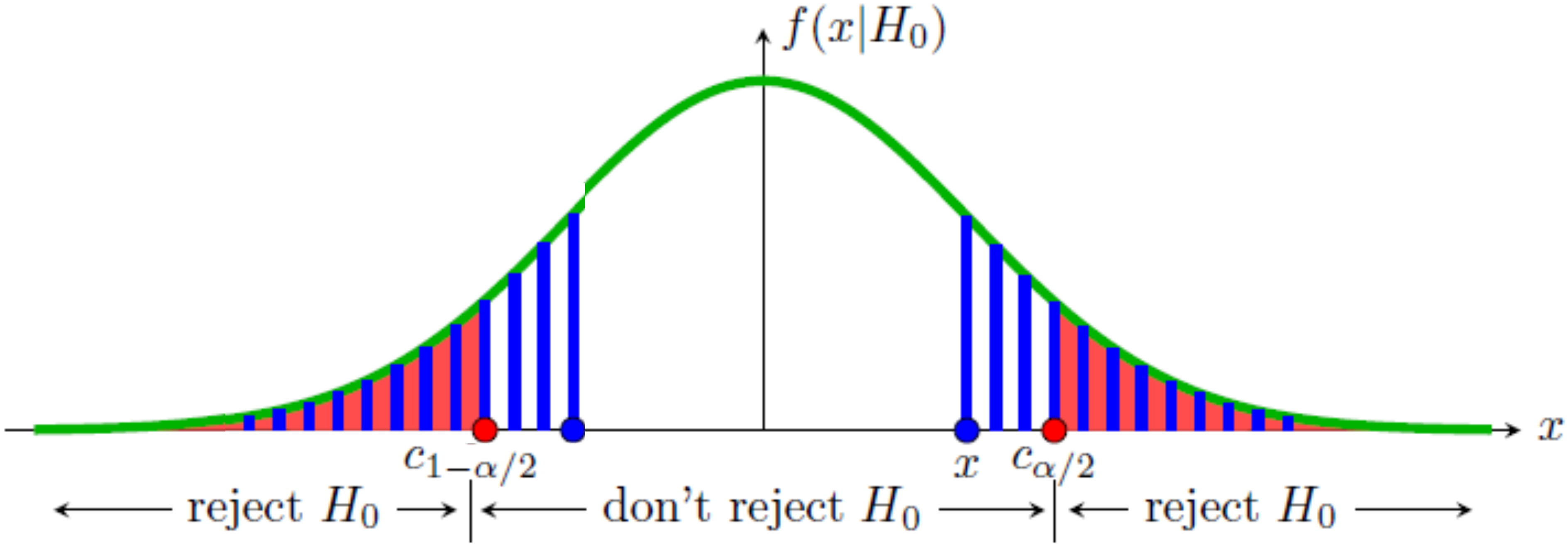
- Two-sided p-values are a bit trickier:
  - What does ‘data at least as extreme’ means in this context?



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): $p = P(T \leq t)$
- Two-sided alternative: $p = 2\min\{P(T > t), P(T \leq t)\}$

$$p < \alpha \implies \textbf{Reject } H_0$$

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.):

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

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

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

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**.

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

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: $n = 775, \quad \bar{X} = 39.08, \quad s_X^2 = 7.77$
  - Emergency: $m = 633, \quad \bar{Y} = 39.60, \quad s_Y^2 = 4.95$

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

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: $m = 633, \;\; \bar{Y} = 39.60, \;\; s_Y^2 = 4.95$

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

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

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.):

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

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

- 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: $H_0\colon \mu_X - \mu_Y = 0$
- 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 $(X_i, Y_i), i = 1, \dots, n$.
- "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?*

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?*
- Samples are paired = dependent.

PAIRED SAMPLE T-TEST

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

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

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

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 **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: $D \sim N(\mu, \sigma^2)$

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

PAIRED SAMPLE T-TEST

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

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

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

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

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

- 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:

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

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

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:

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

NON-PARAMETRIC TESTS

NON-PARAMETRIC TESTS

- Up till now: **parametric** models
  - Example: mean μ and variance σ^2

NON-PARAMETRIC TESTS

- Up till now: **parametric** models
  - Example: mean μ and variance σ^2

- **Non-parametric statistics** doesn’t rely on data belonging to any parametric family of probability distributions:
  - distribution-free or
  - having a specified distribution but with the distribution's parameters unspecified.

NON-PARAMETRIC TESTS

PARAMETRIC HYPOTHESES	NON-PARAMETRIC HYPOTHESES
• Data comes from the normal distribution with specified mean and variance.	
• Data comes from the normal distribution with specified mean and unspecified variance.	

NON-PARAMETRIC TESTS

PARAMETRIC HYPOTHESES	NON-PARAMETRIC HYPOTHESES
• Data comes from the normal distribution with specified mean and variance.	• Data comes from a normal distribution form with both mean and variance unspecified.
• Data comes from the normal distribution with specified mean and unspecified variance.	• Two unspecified continuous distributions are identical.

MANN-WHITNEY U TEST

An alternative to the two-sample t-test
when the distribution of the data cannot be assumed to be normal

MANN-WHITNEY U TEST

- Two independent i.i.d. sets of data:

$$X_1, X_2, \dots, X_n$$

$$Y_1, Y_2, \dots, Y_m$$

MANN-WHITNEY U TEST

- Two independent i.i.d. sets of data:

$$X_1, X_2, \dots, X_n$$

$$Y_1, Y_2, \dots, Y_m$$

- Null hypothesis:

H_0: For randomly selected values X and Y from two populations,

$$P(X > Y) = P(X < Y)$$

MANN-WHITNEY U TEST

- Two independent i.i.d. sets of data:

$$X_1, X_2, \dots, X_n$$

$$Y_1, Y_2, \dots, Y_m$$

- Null hypothesis:

H_0: For randomly selected values *X* and *Y* from two populations,

$$P(X > Y) = P(X < Y)$$

- Alternatives
  - Two-sided: $P(X > Y) \neq P(X < Y)$
  - One-sided: $P(X > Y) > P(X < Y)$, $\quad P(X > Y) < P(X < Y)$

MANN-WHITNEY U TEST

- Test statistic:

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} S(X_i, Y_j)$$

where $S(X,Y) = \begin{cases} 1, & Y < X \\ 1/2, & Y = X \\ 0, & Y > X \end{cases}$

MANN-WHITNEY U TEST

- Test statistic:

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} S(X_i, Y_j)$$

$$\text{where } S(X,Y) = \begin{cases} 1, & Y < X \\ 1/2, & Y = X \\ 0, & Y > X \end{cases}$$

- Null distribution:

For small n, tabulated

For large n, $(n \geq 30)$ $U \sim \text{normal}$

WILCOXON SIGNED RANK TEST

An alternative to the paired t-test
when the distribution of the differences cannot be assumed to be normal

WILCOXON SIGNED RANK TEST

- Two paired i.i.d. sets of data:

$$X_1, X_2, \dots, X_n$$

$$Y_1, Y_2, \dots, Y_m$$

WILCOXON SIGNED RANK TEST

- Two paired i.i.d. sets of data:

$$X_1, X_2, \dots, X_n$$

$$Y_1, Y_2, \dots, Y_m$$

- Null hypothesis:

H_0: difference between the pairs follows a symmetric distribution around zero

WILCOXON SIGNED RANK TEST

- Two paired i.i.d. sets of data:

$$X_1, X_2, \dots, X_n$$

$$Y_1, Y_2, \dots, Y_m$$

- Null hypothesis:

H_0: difference between the pairs follows a symmetric distribution around zero

- Alternatives
  - Two-sided: difference between the pairs doesn't follow a symmetric distribution around zero
  - One-sided: distribution is skewed to one particular side

WILCOXON SIGNED RANK TEST

- Test statistic:

WILCOXON SIGNED RANK TEST

- Test statistic:

$$\text{Compute } |X_i - Y_i| \text{ and } \operatorname{sgn}(X_i - Y_i)$$

WILCOXON SIGNED RANK TEST

- Test statistic:

Compute $|X_i - Y_i|$ and $\text{sgn}(X_i - Y_i)$

Exclude pairs (X_i, Y_i) where$|X_i - Y_i| = 0$

WILCOXON SIGNED RANK TEST

- Test statistic:

Compute $|X_i - Y_i|$ and $\mathrm{sgn}(X_i - Y_i)$

Exclude pairs (X_i, Y_i) where $|X_i - Y_i| = 0$

Rank the remaining pairs from smallest to largest $|X_i - Y_i| \rightarrow R_i$

WILCOXON SIGNED RANK TEST

- Test statistic:

Compute $|X_i - Y_i|$ and $\mathrm{sgn}(X_i - Y_i)$

Exclude pairs (X_i, Y_i) where $|X_i - Y_i| = 0$

Rank the remaining pairs from smallest to largest $|X_i - Y_i| \rightarrow R_i$

$$W = \sum_{i=1}^{n} R_i \, \mathrm{sgn}(X_i - Y_i)$$

WILCOXON SIGNED RANK TEST

- Test statistic:

Compute $|X_i - Y_i|$ and $\text{sgn}(X_i - Y_i)$

Exclude pairs (X_i, Y_i) where $|X_i - Y_i| = 0$

Rank the remaining pairs from smallest to largest $|X_i - Y_i| \rightarrow R_i$

$$W = \sum_{i=1}^{n} R_i \, \text{sgn}(X_i - Y_i)$$

- Null distributions:
  - some specific tabulated distribution.

χ^2-TEST

Independence of two categorical variables based on contingency table

χ^2-TEST

- Motivating example: consider the following **contingency table**:

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

χ^2-TEST

- Motivating example: consider the following **contingency table**:

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

- Are education levels and number of marriages (one / many) independent?

χ^2-TEST

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

- Are education levels and number of marriages (one / many) independent?

If so, cell probabilities are the product of the marginal ones:

Education	Married once	Married multiple times	Total
College			$611/1436$
No college			$825/1436$
Total	$1231/1436$	$205/1436$	1

χ^2-TEST

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

- Are education levels and number of marriages (one / many) independent?

If so, cell probabilities are the product of the marginal ones:

Education	Married once	Married multiple times	Total
College	0.365		$611/1436$
No college			$825/1436$
Total	$1231/1436$	$205/1436$	1

χ^2-TEST

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

- Are education levels and number of marriages (one / many) independent?

If so, cell probabilities are the product of the marginal ones:

Education	Married once	Married multiple times	Total
College	0.365	0.061	$611/1436$
No college			$825/1436$
Total	$1231/1436$	$205/1436$	1

χ^2-TEST

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

- Are education levels and number of marriages (one / many) independent?

If so, cell probabilities are the product of the marginal ones:

Education	Married once	Married multiple times	Total
College	0.365	0.061	$611/1436$
No college	0.492		$825/1436$
Total	$1231/1436$	$205/1436$	1

χ^2-TEST

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

- Are education levels and number of marriages (one / many) independent?

If so, cell probabilities are the product of the marginal ones:

Education	Married once	Married multiple times	Total
College	0.365	0.061	611/1436
No college	0.492	0.082	825/1436
Total	1231/1436	205/1436	1

χ^2-TEST

Education	Married once	Married multiple times	Total
College	0.365	0.061	611/1436
No college	0.492	0.082	825/1436
Total	1231/1436	205/1436	1

- From this, we can get ***expected counts*** *(multiplying cell probabilities by the total number of women surveyed):*

Education	Married once	Married multiple times
College	550, 523.8	61,
No college	681,	144,

χ^2-TEST

Education	Married once	Married multiple times	Total
College	0.365	0.061	$611/1436$
No college	0.492	0.082	$825/1436$
Total	$1231/1436$	$205/1436$	1

- From this, we can get ***expected counts*** *(multiplying cell probabilities by the total number of women surveyed):*

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681,	144,

χ^2-TEST

Education	Married once	Married multiple times	Total
College	0.365	0.061	$611/1436$
No college	0.492	0.082	$825/1436$
Total	$1231/1436$	$205/1436$	1

- From this, we can get ***expected counts*** *(multiplying cell probabilities by the total number of women surveyed):*

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144,

χ^2-TEST

Education	Married once	Married multiple times	Total
College	0.365	0.061	$611/1436$
No college	0.492	0.082	$825/1436$
Total	$1231/1436$	$205/1436$	1

- From this, we can get ***expected counts*** *(multiplying cell probabilities by the total number of women surveyed):*

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- Null hypothesis:

H_0: cell probabilities are the product of the marginal ones, the difference, so the difference between them should be small

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- Null hypothesis:

H_0: cell probabilities are the product of the marginal ones, the difference, so the difference between them should be small

H_1: the difference between observed and expected counts is large **(!!! one-sided)**

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- Test statistic: Pearson's chi-square statistic

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(df), \qquad df = (n-1)(m-1)$$

where O_i - observed count, and E_i - expected count

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- In this case

$$\chi^2 =$$

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- In this case

$$\chi^2 = 16.01, \ df = 1$$

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- In this case

$$\chi^2 = 16.01, \ df = 1$$

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- In this case

$$\chi^2 = 16.01, \;\; df \; = 1$$

$$\chi^2_{1-\alpha}(1) = \chi^2_{0.95}(1) = 7.879$$

χ^2-TEST

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

- In this case

$$\chi^2 = 16.01,\ df\ = 1$$

$$\chi^2_{1-\alpha}(1) = \chi^2_{0.95}(1) = 7.879$$

$$16.01 > 7.879 \Longrightarrow \text{reject } H_0$$

NON-PARAMETRIC TESTS

- Fewer assumptions, wider applicability.

NON-PARAMETRIC TESTS

- Fewer assumptions, wider applicability.

- This comes with the cost: when parametric tests are applicable, non-parametric ones have less *power*
  - *a larger sample size can be required to draw conclusions with the same degree of confidence.*