Copyright Notice

DeepLearning.AI

Math for Machine Learning

-

Probability and Statistics - Week 4

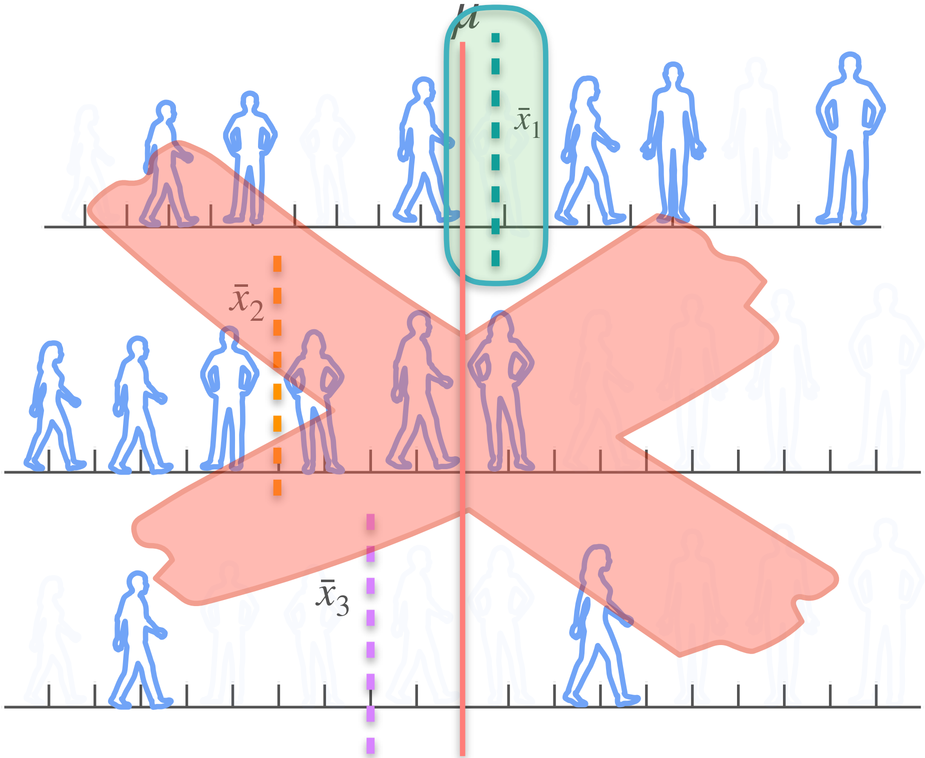
W4 Lesson 1

DeepLearning.AI

Confidence Interval

Confidence Interval (Known Standard Deviation)

Confidence Interval - Intuition



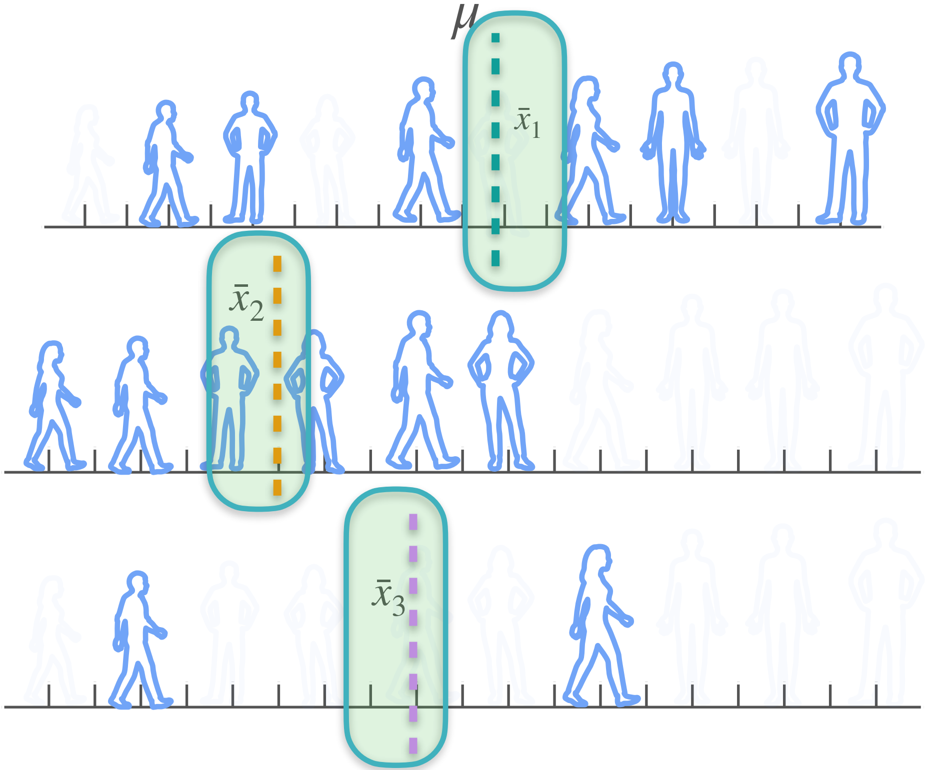
Statistopia

10,000 people

μ
(mean height of the population)

Can you use these sample means with some degree of certainty?

Confidence Interval - Intuition



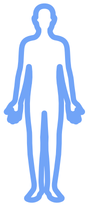
Can you use these sample means with some degree of certainty?

Confidence Interval

lower limit $< \mu <$ upper limit

Confidence Interval - Intuition

$n = 1$



$\bar{x}$

Confidence level

%

$1 - \alpha$

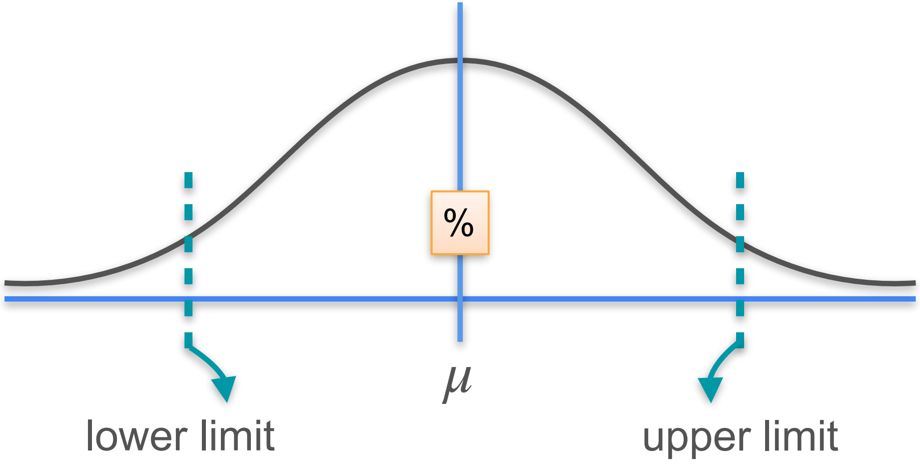
α

significance level

Known

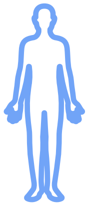
$X \sim \mathcal{N}(\mu, \sigma^2)$

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2)$



Confidence Interval - Intuition

$n = 1$



$\bar{x}$

Confidence level

95%

$1 - \alpha$

$1 - 0.05 = 0.95$

$\alpha = 0.05$

significance level

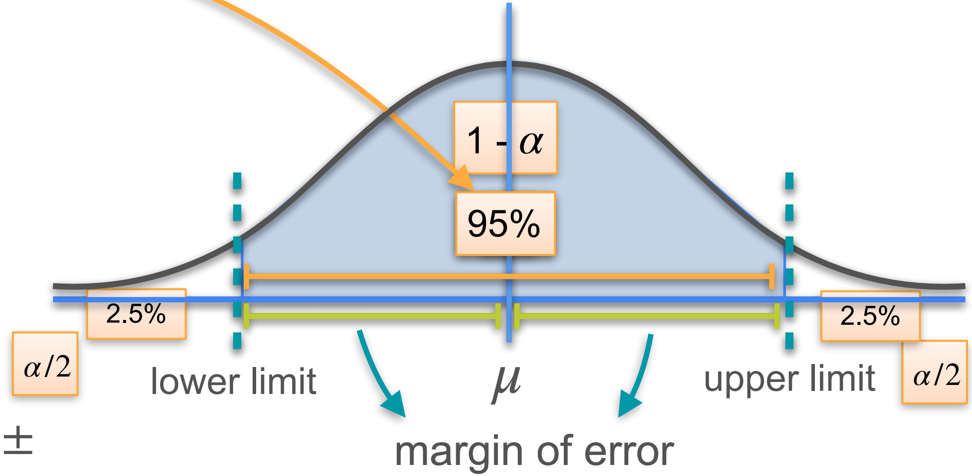
Confidence Interval

$\pm$

Known

$X \sim \mathcal{N}(\mu, \sigma^2)$

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2)$



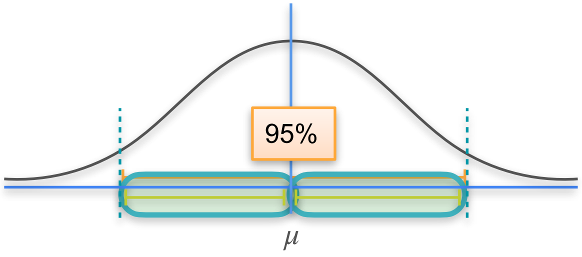
Confidence Interval - Intuition

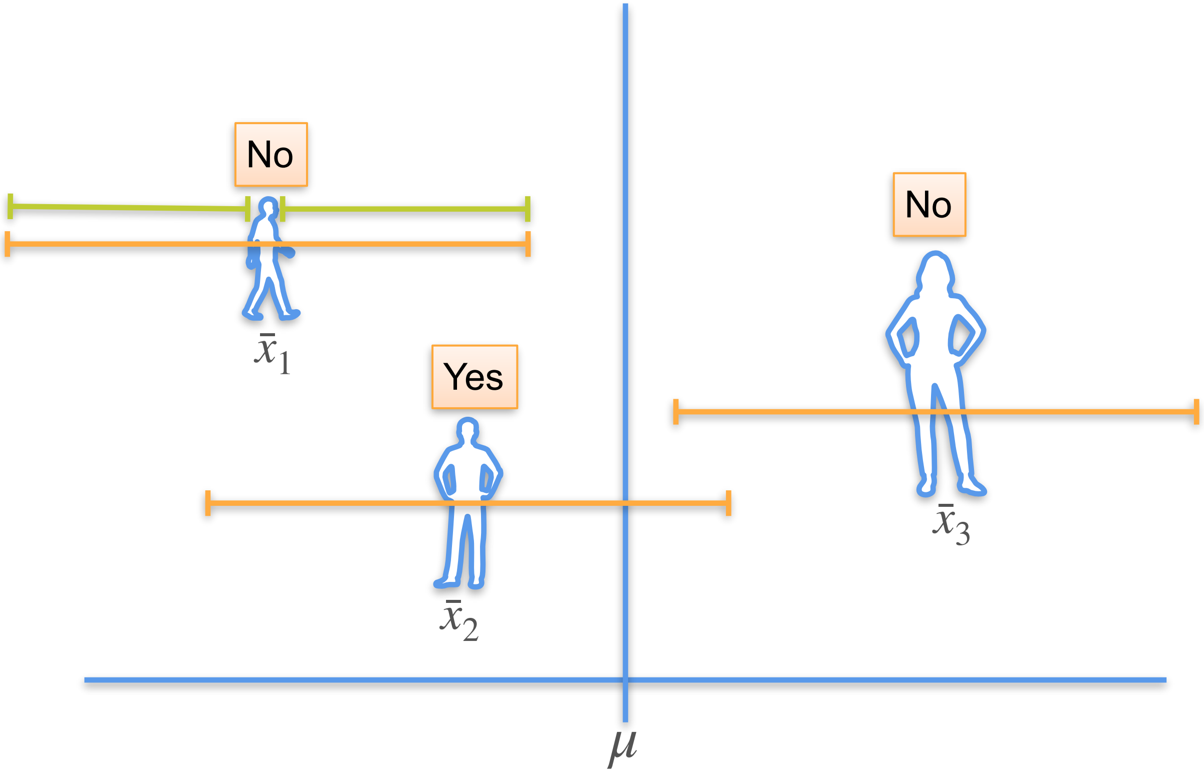
$$n = 1$$

Known σ

95%

Margin of error





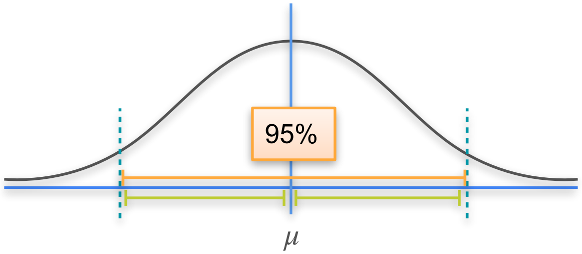
Confidence Interval - Intuition

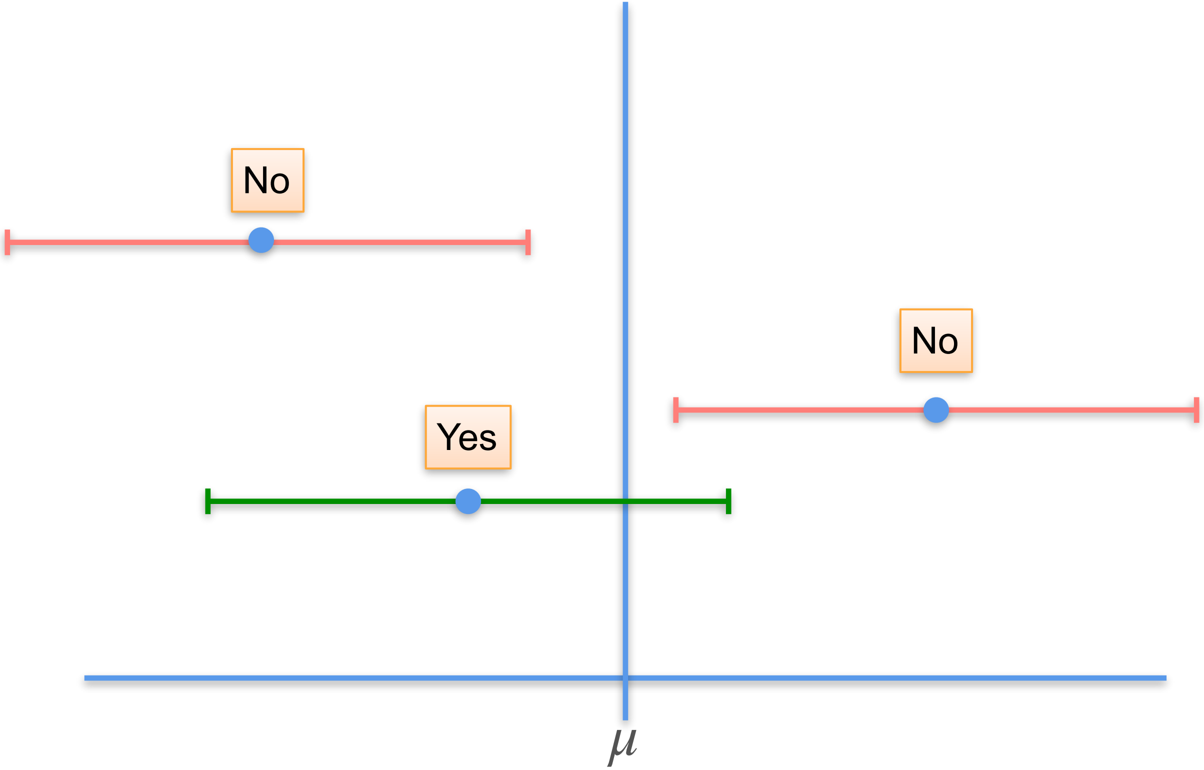
$n = 1$

Known σ

95%

Margin of error





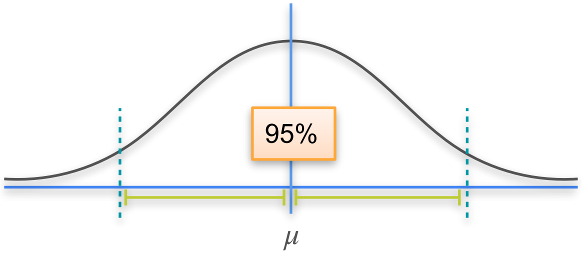
Confidence Interval - Intuition

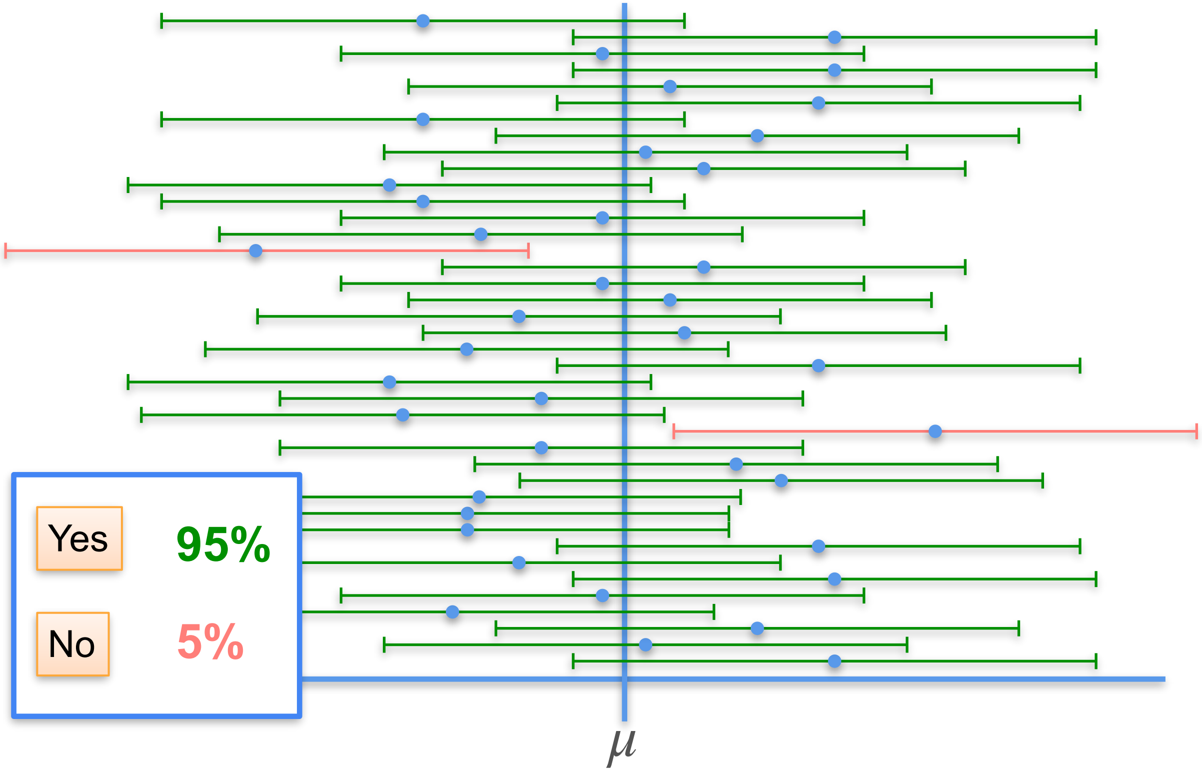
$$n = 1$$

Known σ

95%

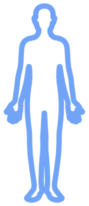
Margin of error





Confidence Interval - Intuition

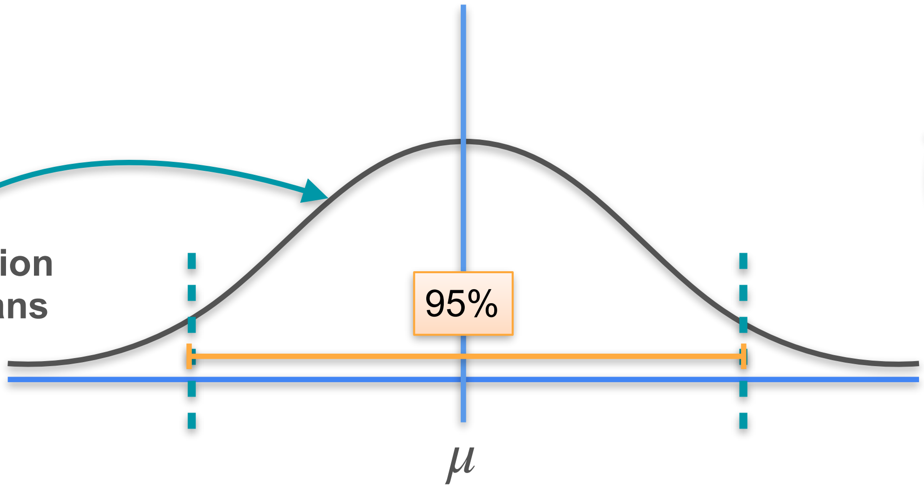
$n = 1$



$\bar{x}$

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2)$$

sampling distribution of the sample means



mean of the sample means

$$\mu_{\bar{x}} = \mu$$

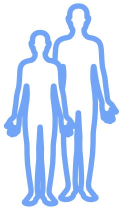
population mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{1}} = \sigma$$

standard deviation of the sample means

Confidence Interval - Intuition

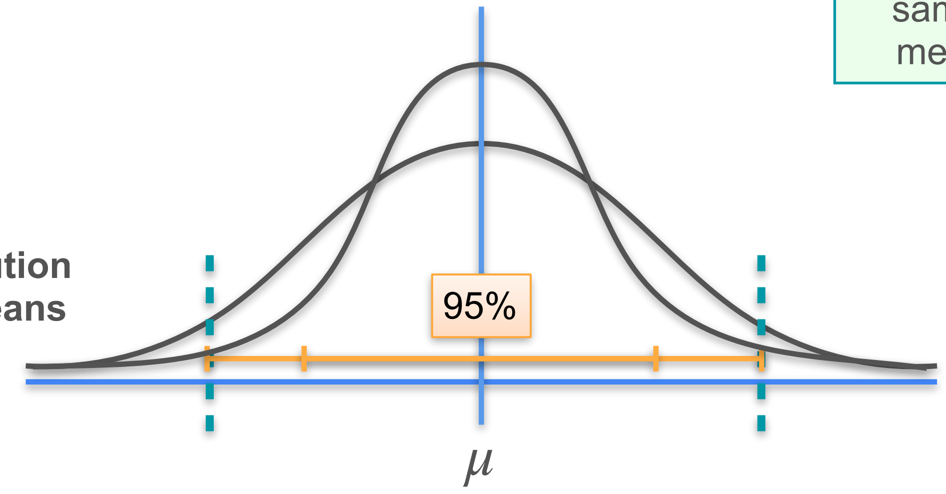
$$n = 2$$



$\bar{x}$

sampling distribution of the sample means

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2}\right)$$



Known population standard deviation (σ)

mean of the sample means

$$\mu_{\bar{x}} = \mu$$

population mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{2}}$$

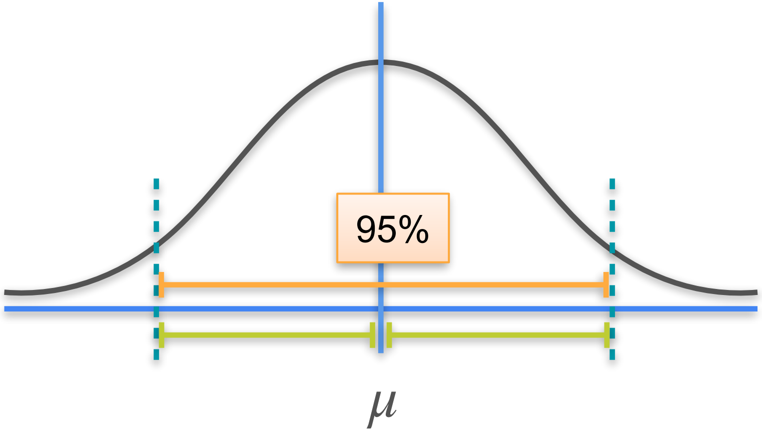
standard deviation of the sample means

Confidence Interval - Intuition

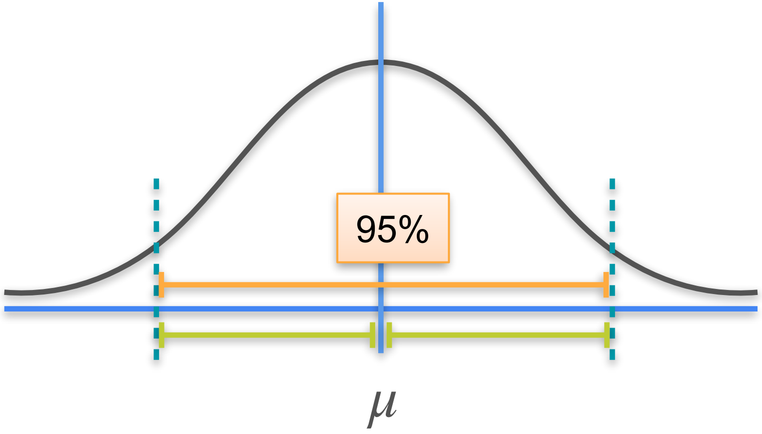
95%

$$n = 1$$

$$\mathcal{N}\left(\mu, \frac{\sigma^2}{1}\right)$$

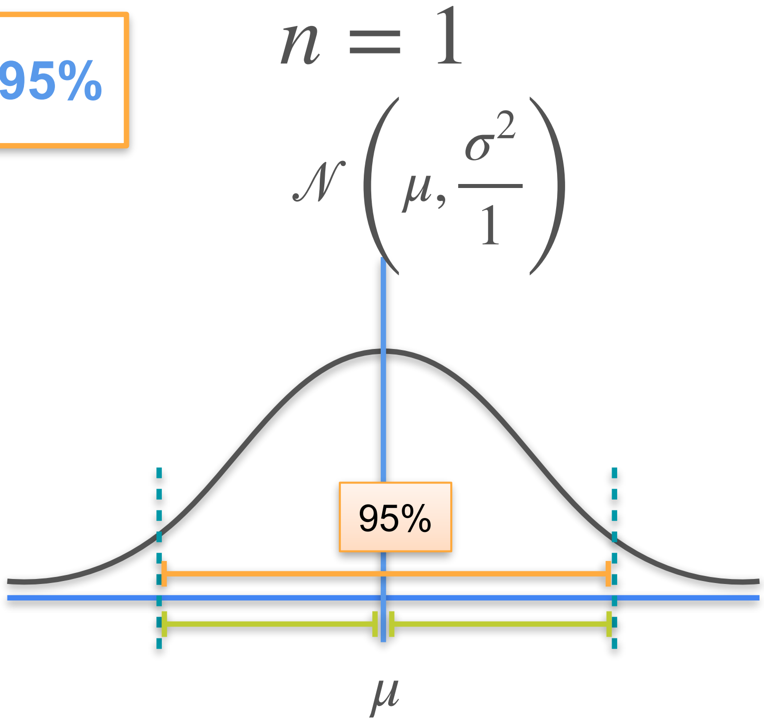


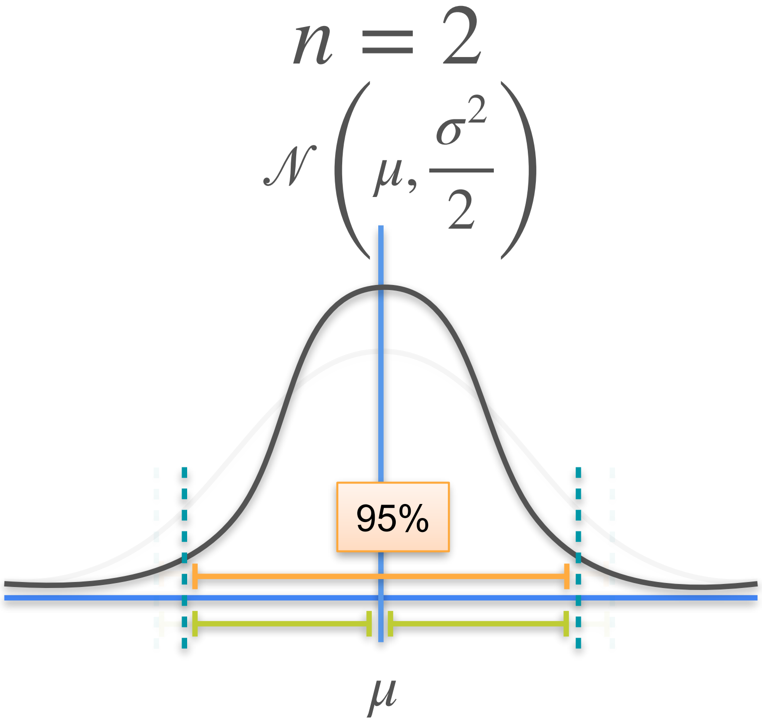
$$n = 2$$



Confidence Interval - Intuition

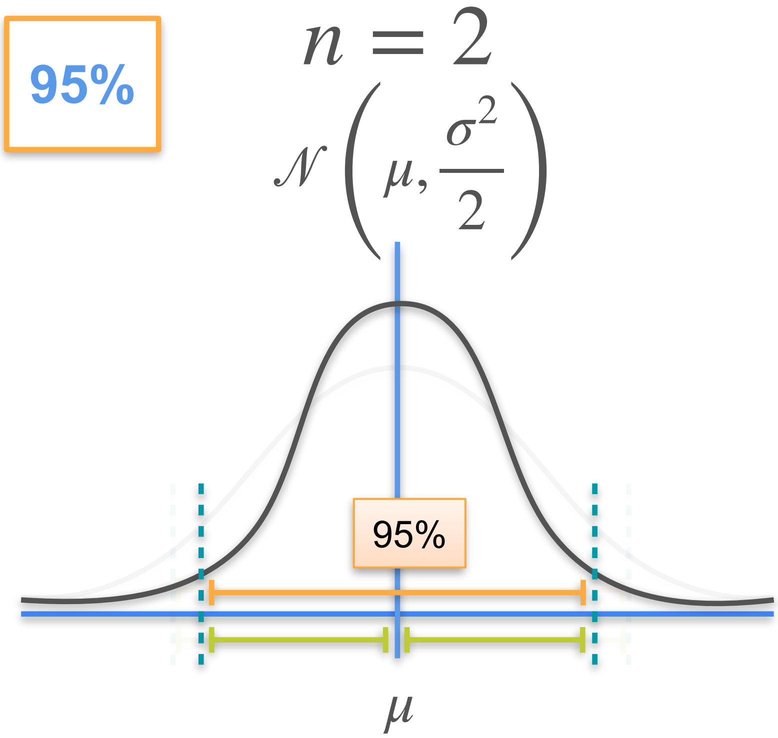
95%

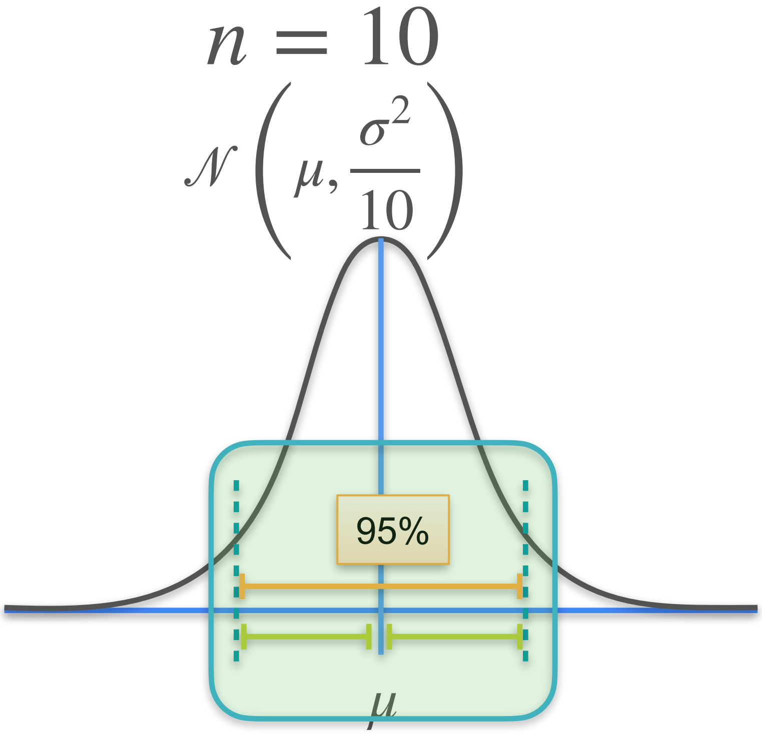




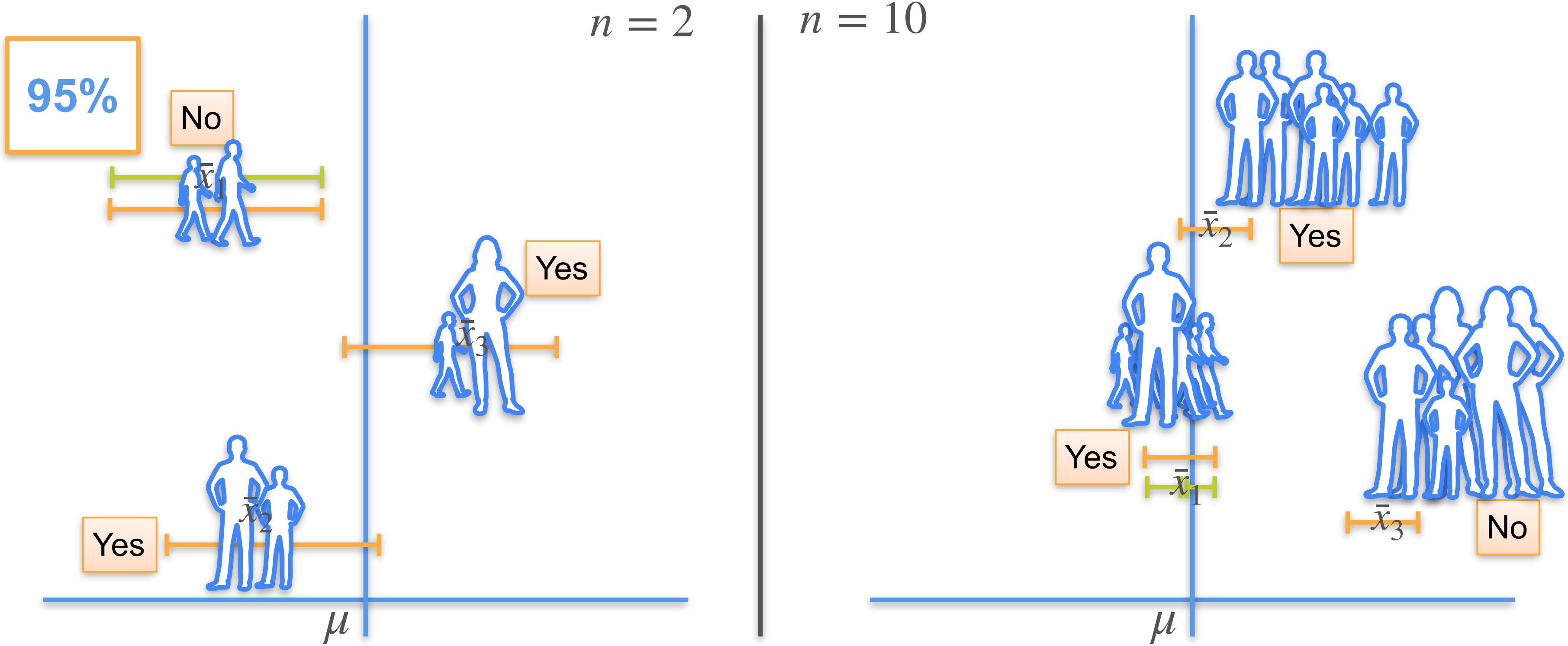
Confidence Interval - Intuition

95%

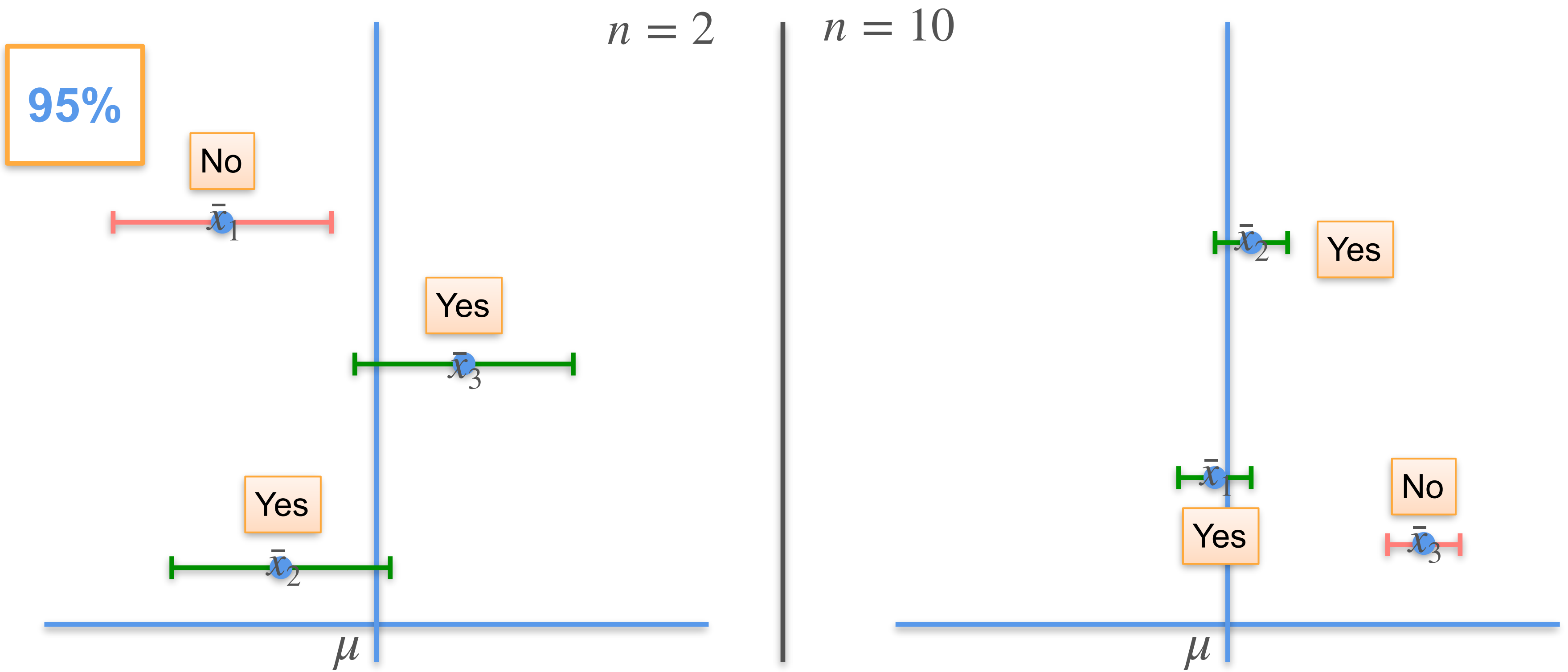




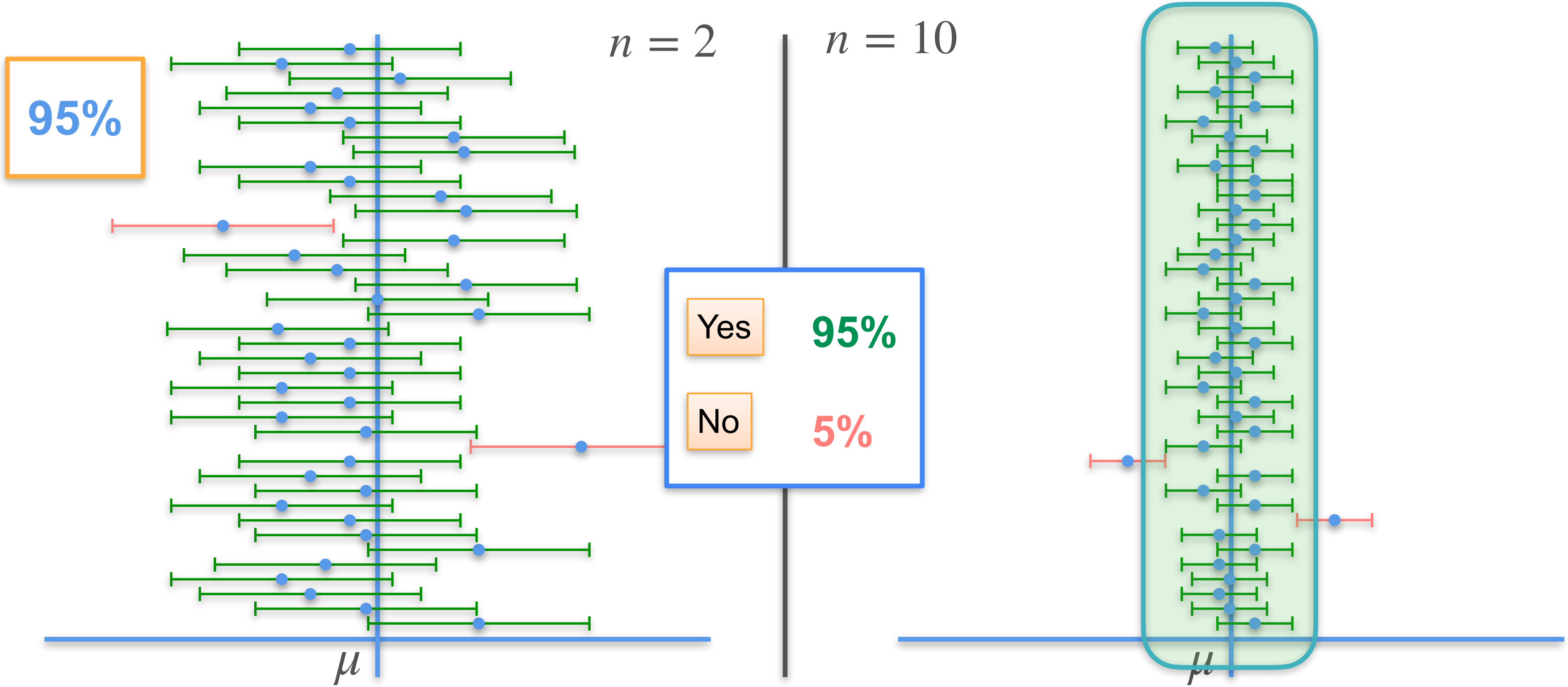
Confidence Interval - Intuition



Confidence Interval - Intuition



Confidence Interval - Intuition



Effect of the Sample Size

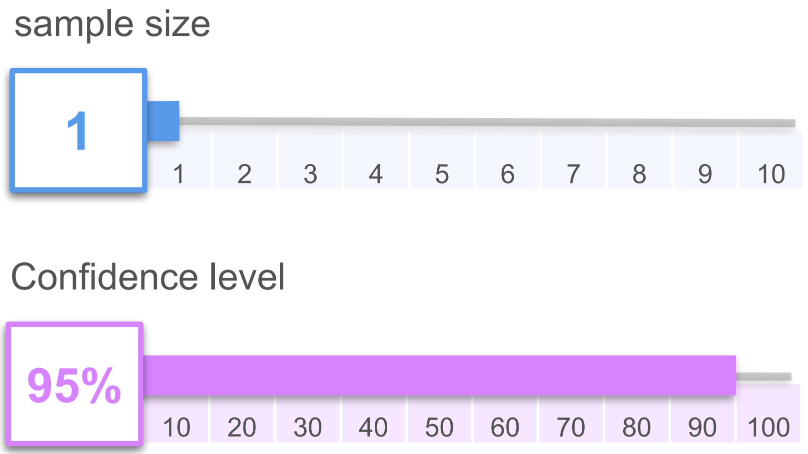
sample size

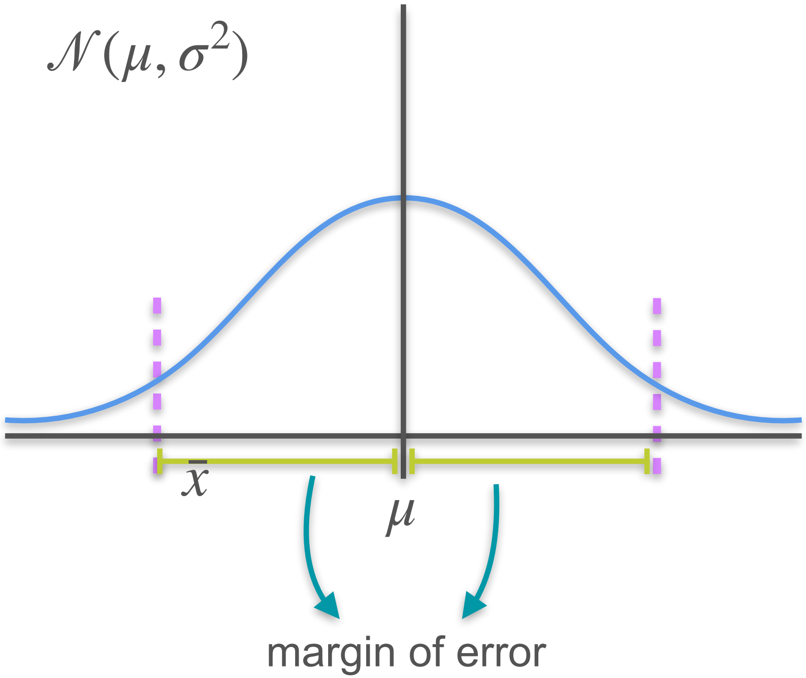
n	1	2	3	4	5	6	7	8	9	10

Confidence level

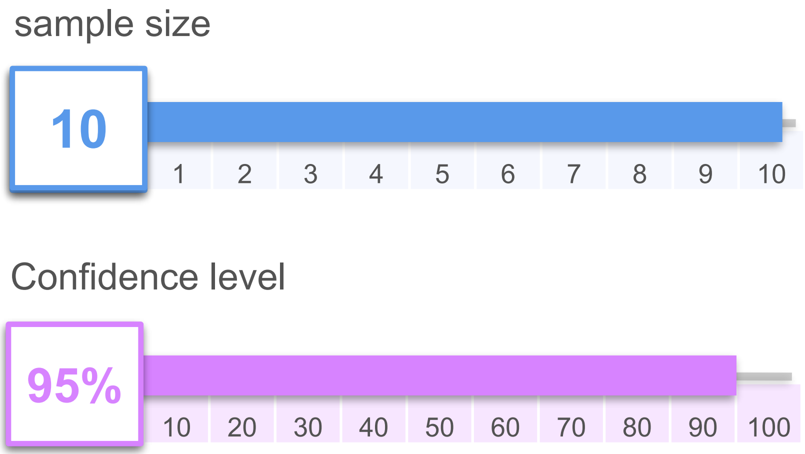
%	10	20	30	40	50	60	70	80	90	100

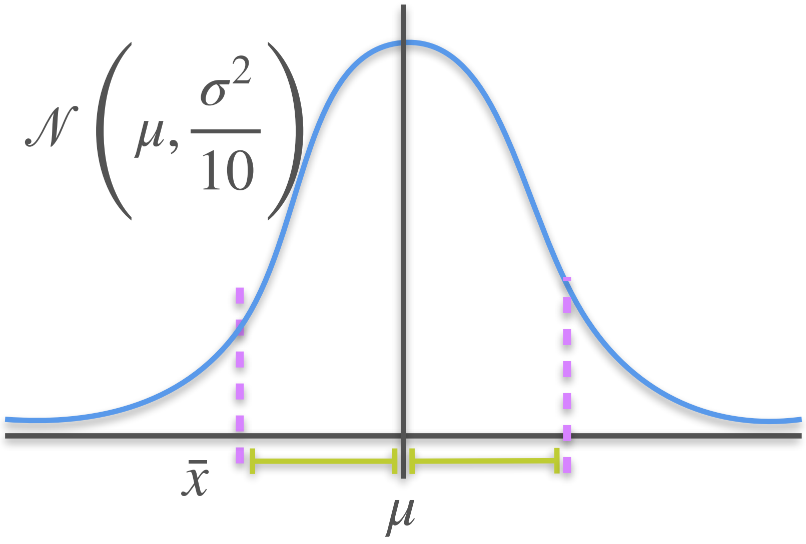
Effect of the Sample Size



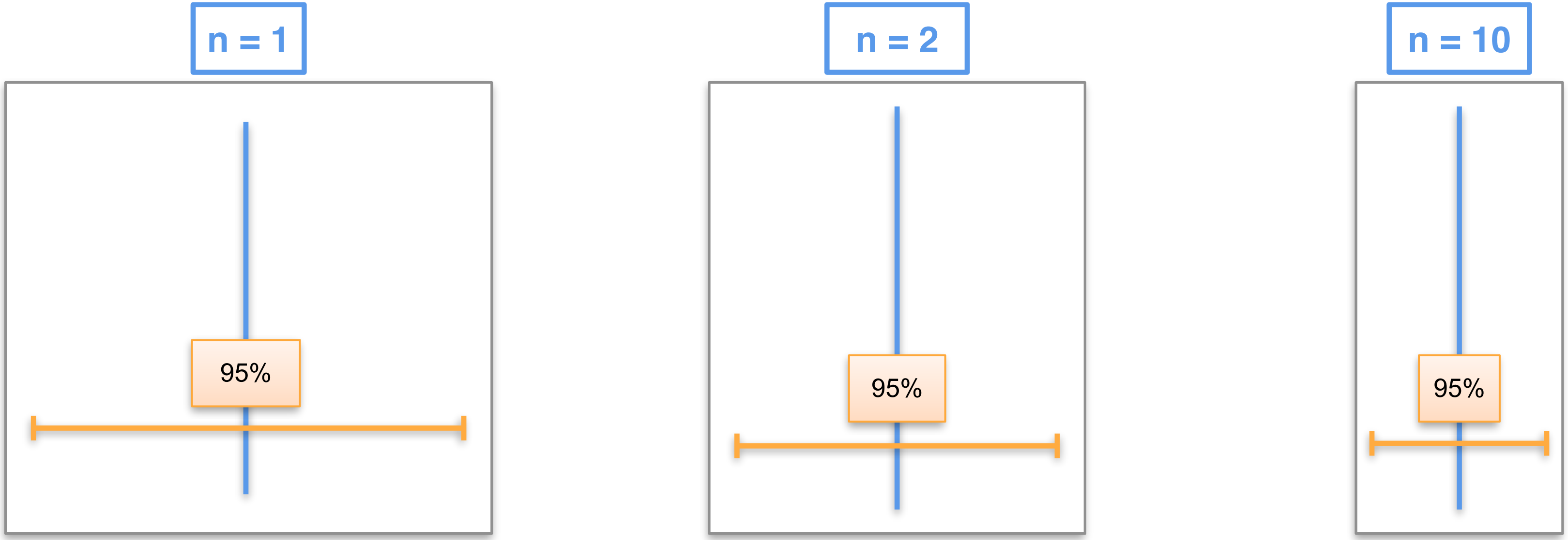


Effect of the Sample Size





Effect of the Sample Size



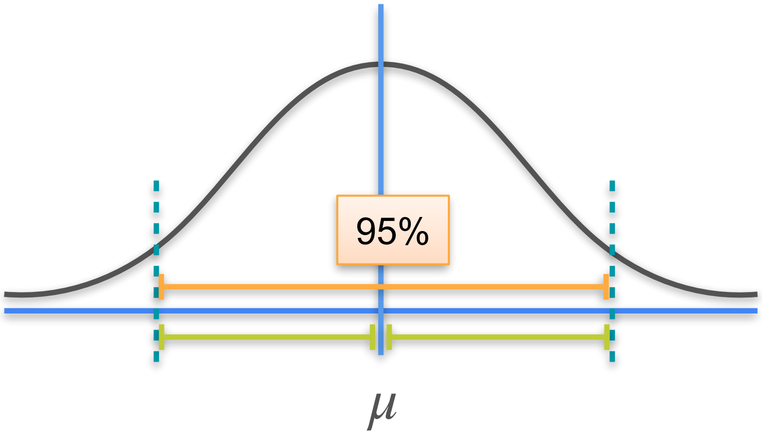
As n increases, the confidence interval shrinks

Effect of the Confidence Level

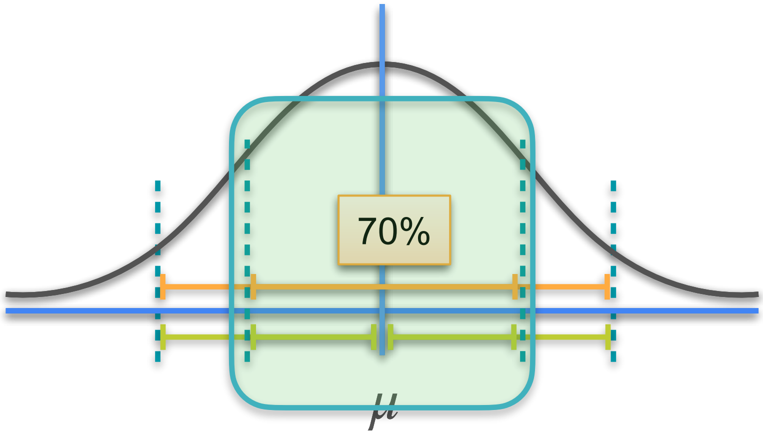
$n = 1$

$\mathcal{N}(\mu, \sigma^2)$

95%

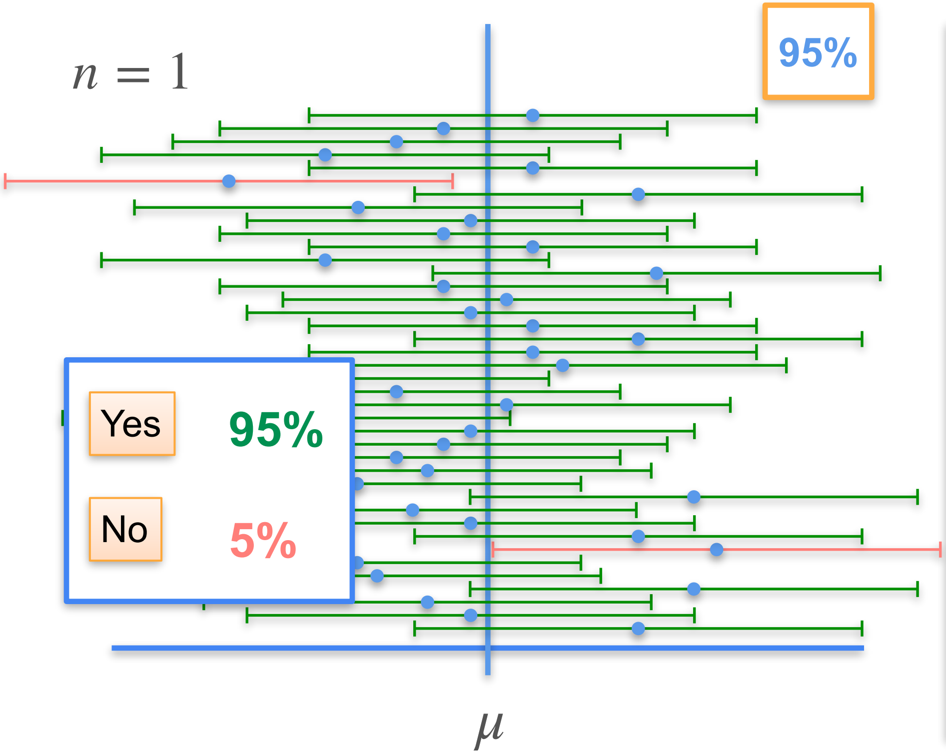


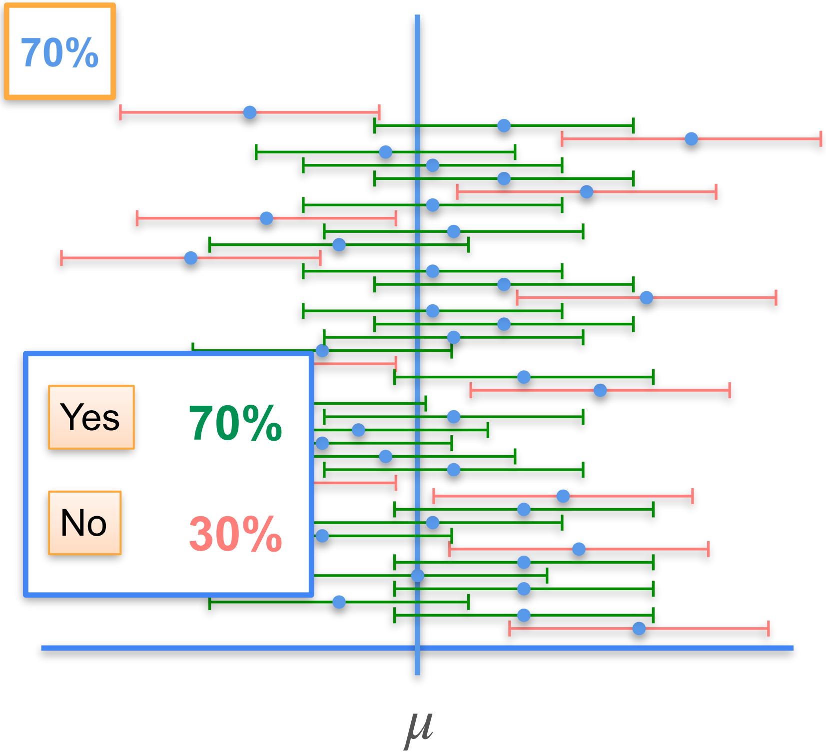
70%



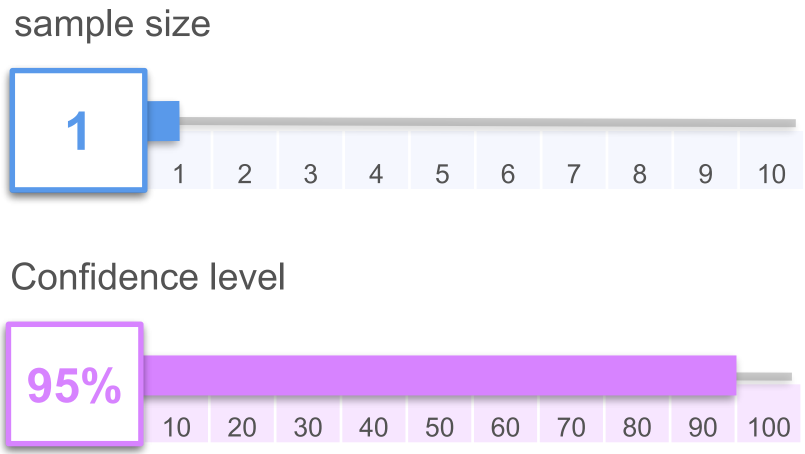
Effect of the Confidence Level

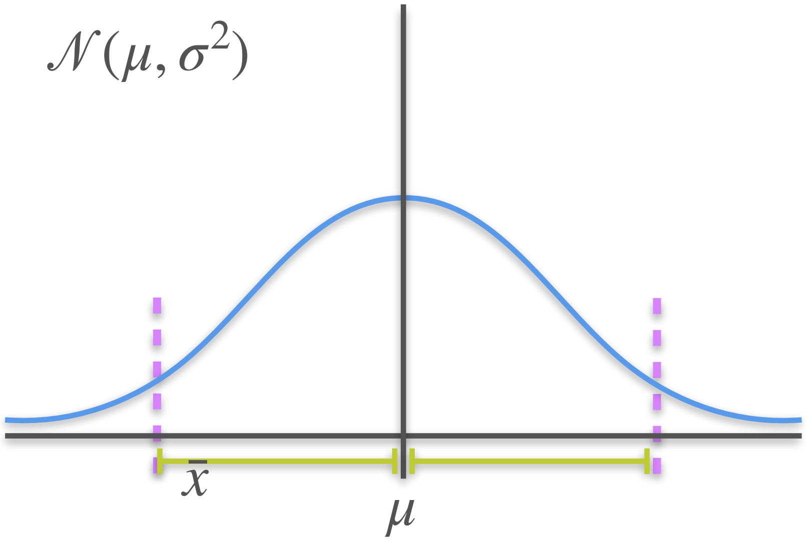
$n = 1$



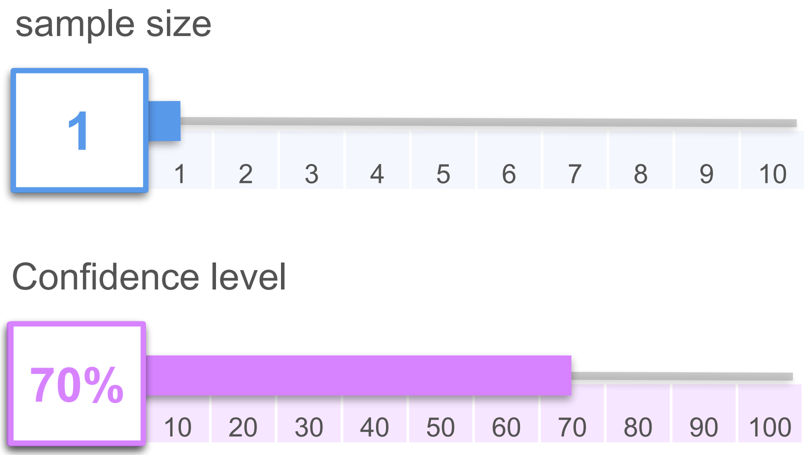


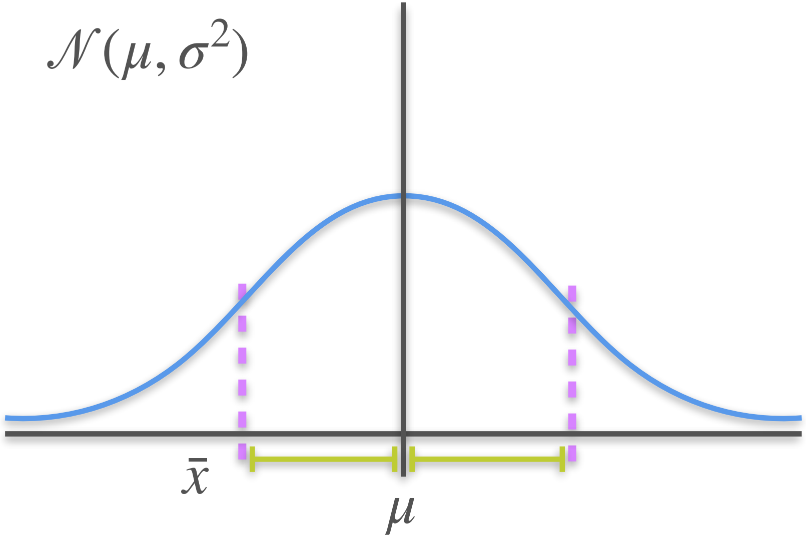
Effect of the Confidence Level



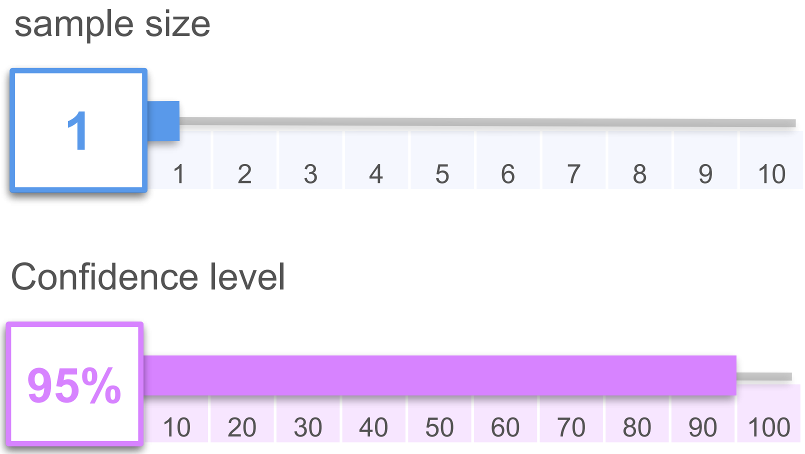


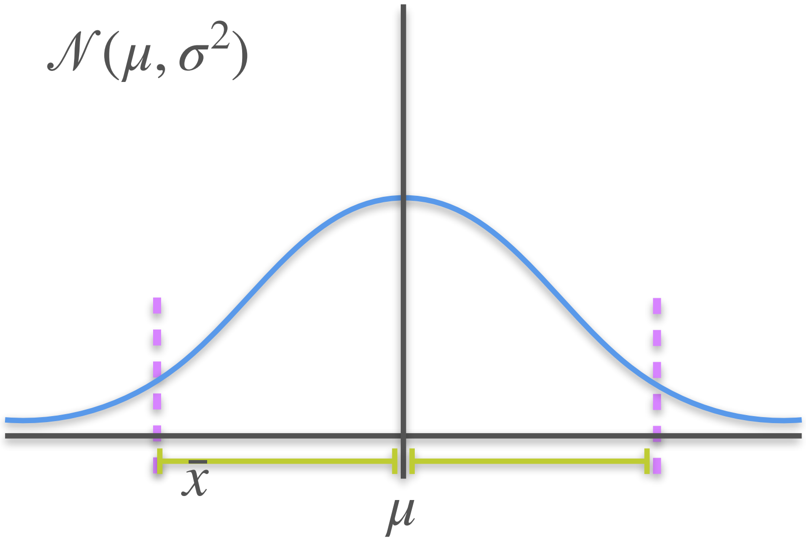
Effect of the Confidence Level





Effect of the Confidence Level



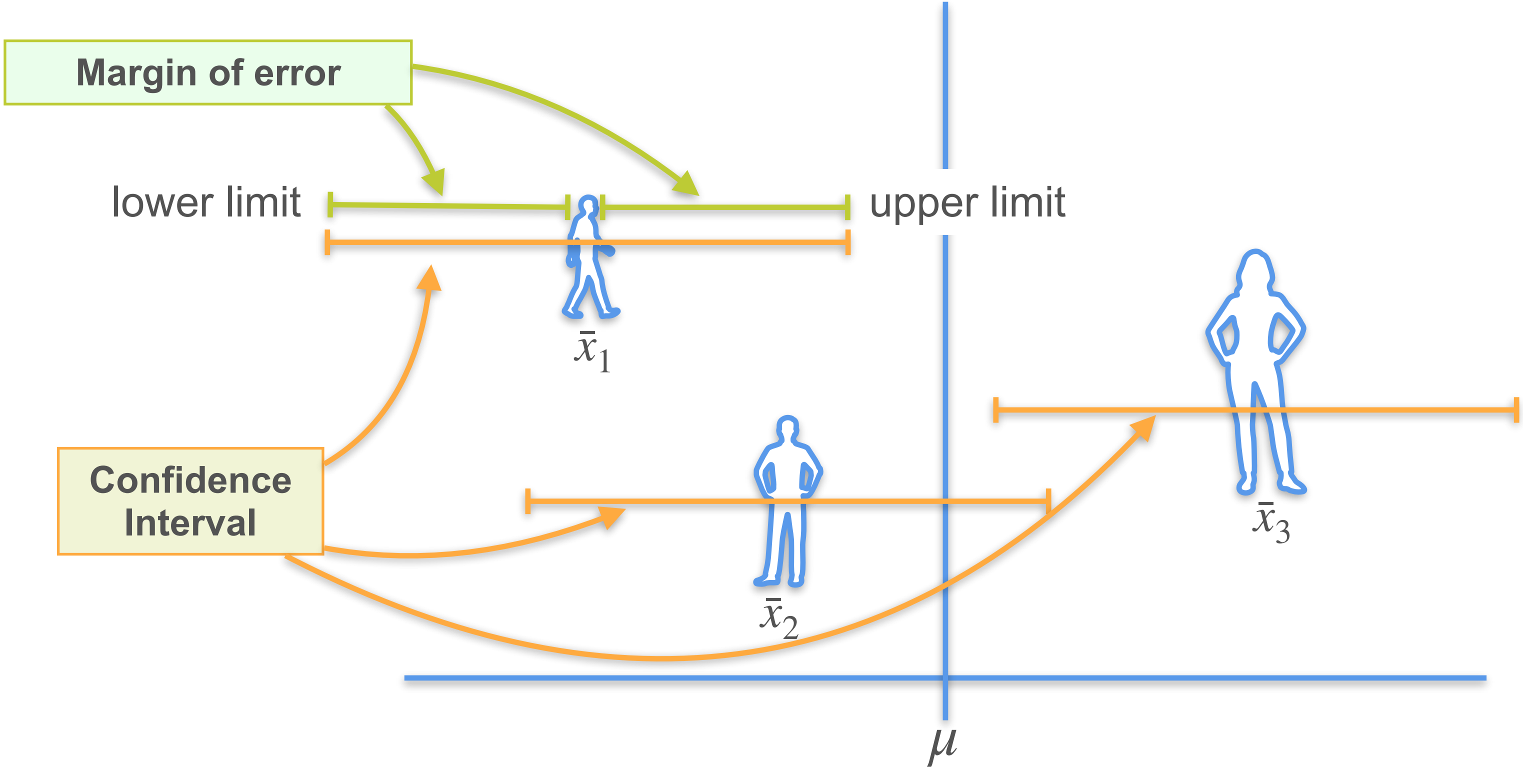


DeepLearning.AI

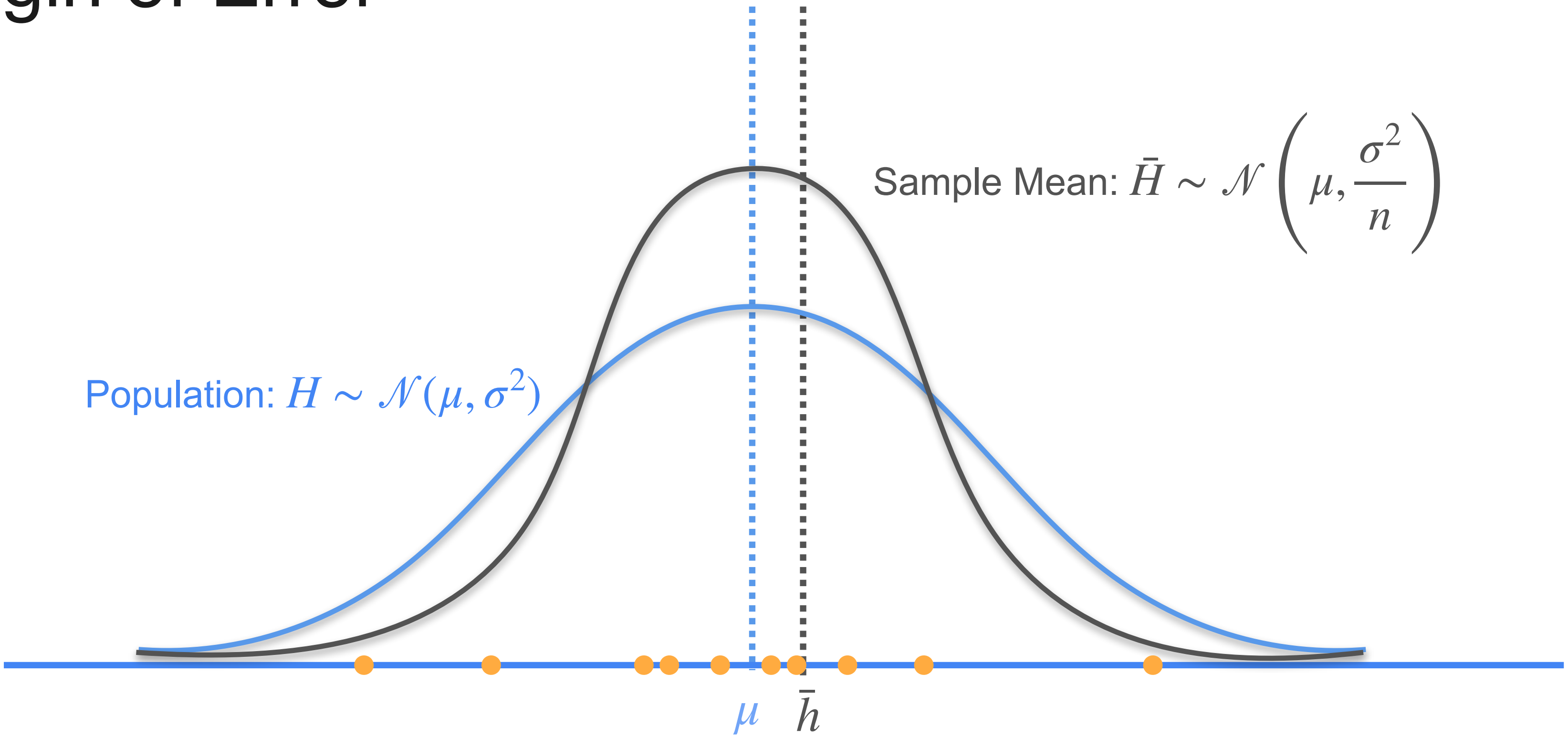
Confidence Interval

Margin of Error

Margin of Error - Introduction

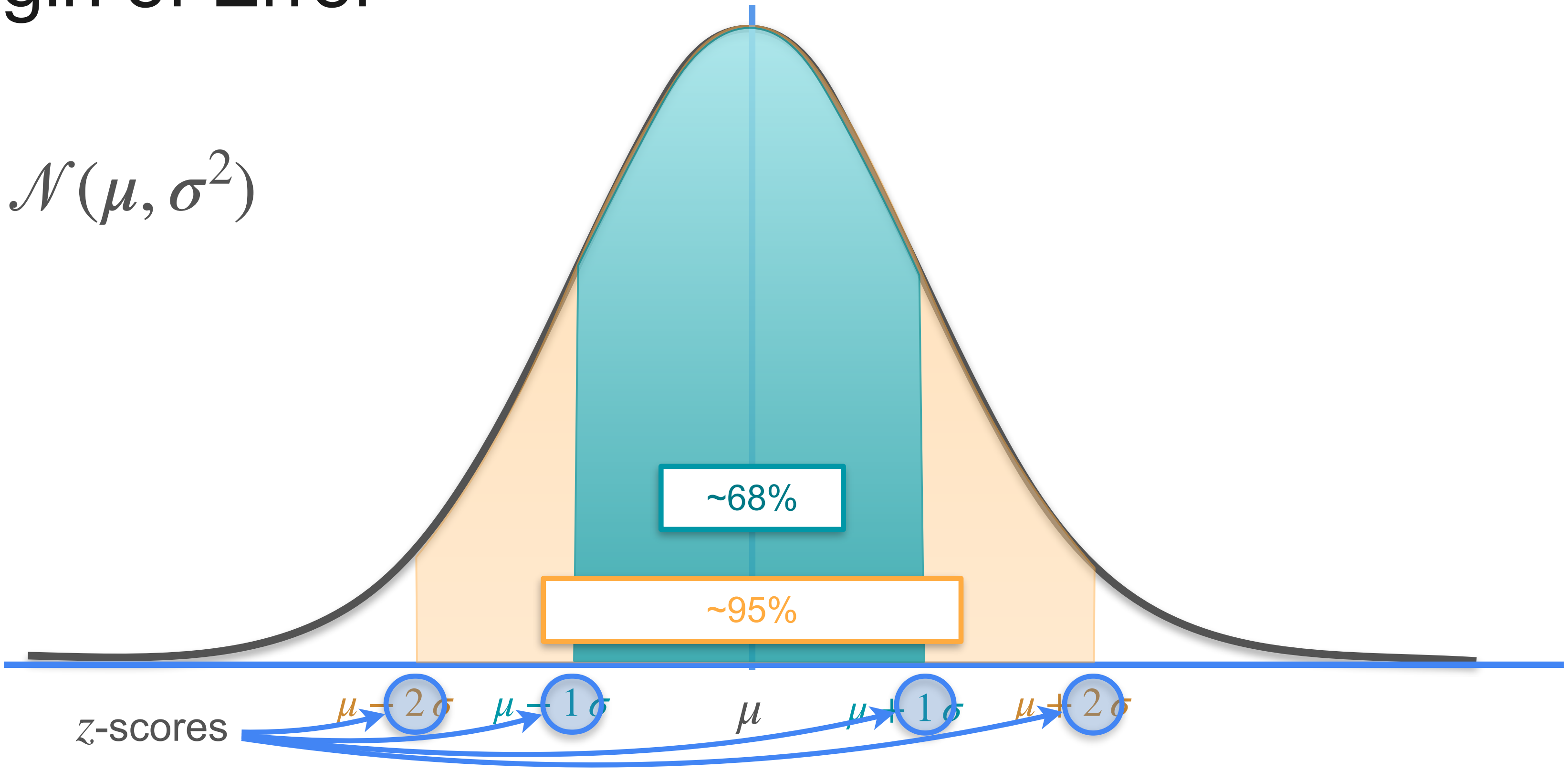


Margin of Error



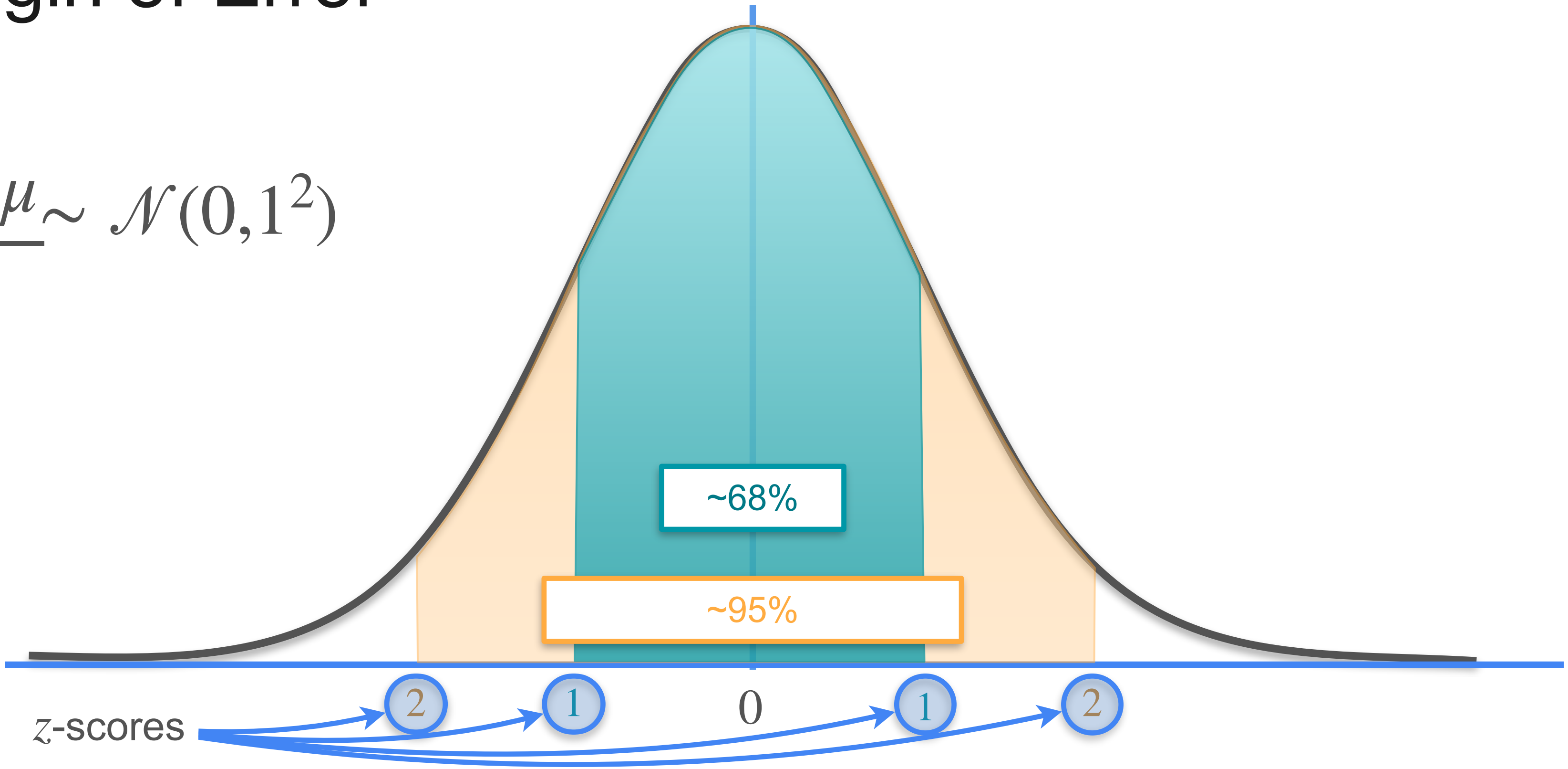
Margin of Error

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



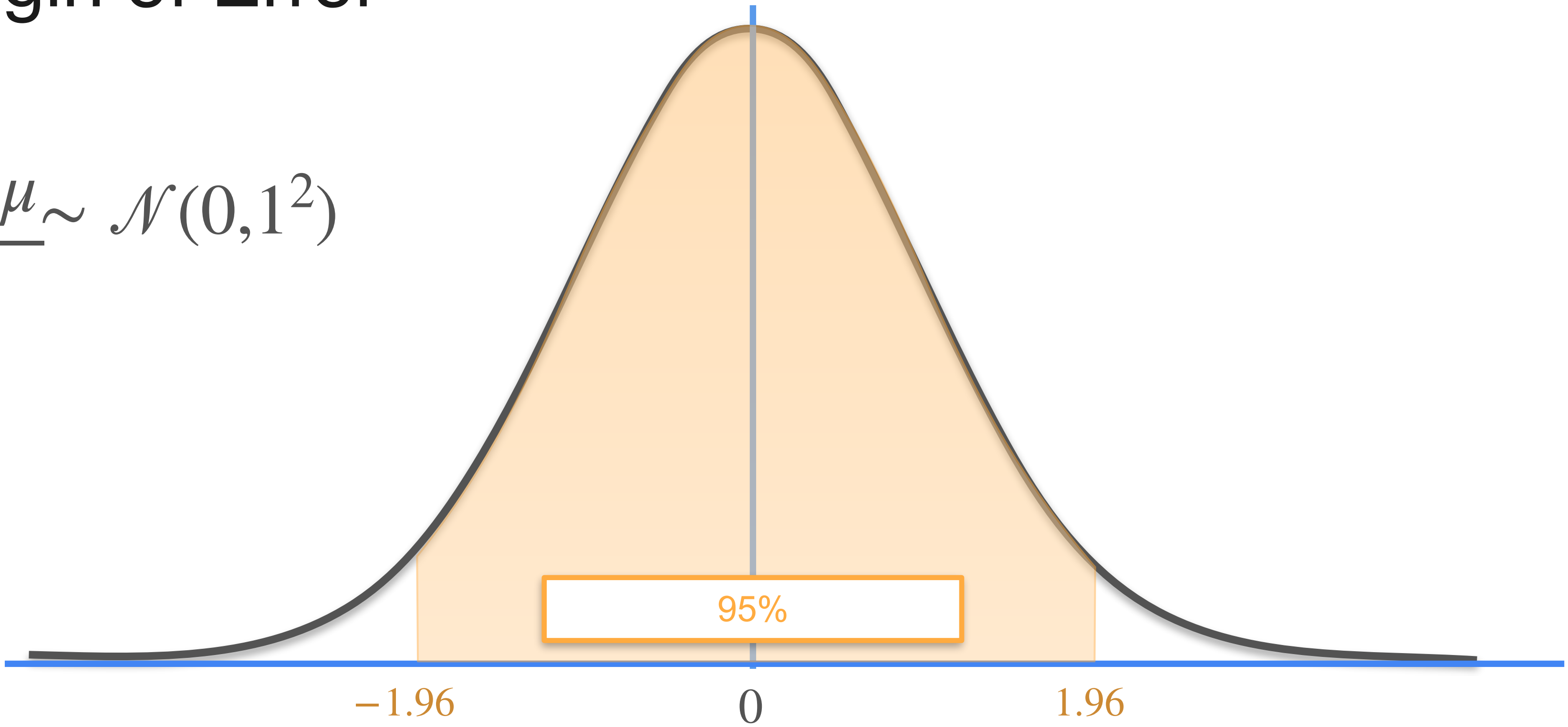
Margin of Error

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1^2)$$



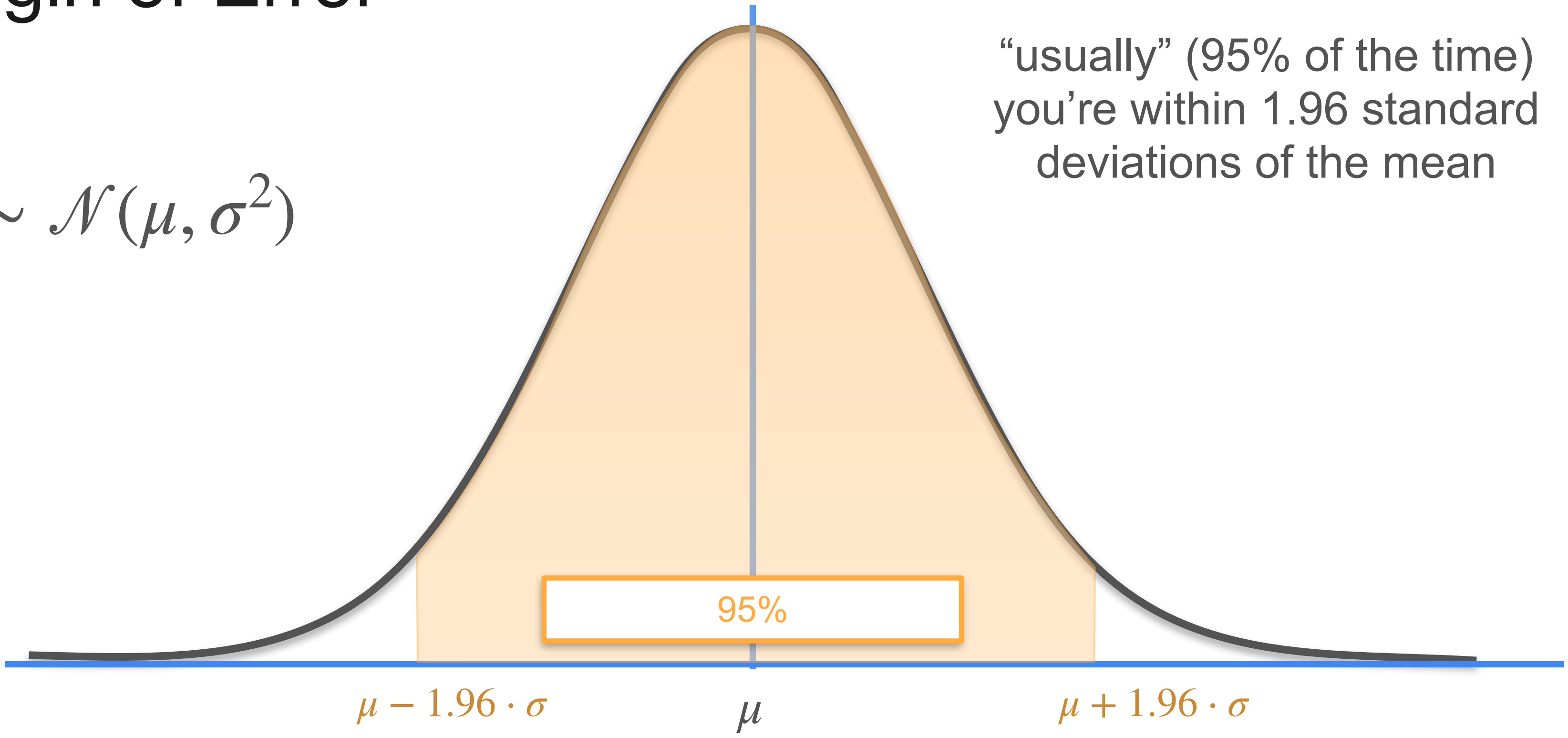
Margin of Error

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1^2)$$



Margin of Error

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



Margin of Error

$$\frac{\bar{X} - \mu}{\sigma} \sim \mathcal{N}(0, 1^2)$$

$$\bar{X} \sim \mathcal{N}\left(\mu_{\bar{x}}, \frac{\sigma^2}{n}\right)$$

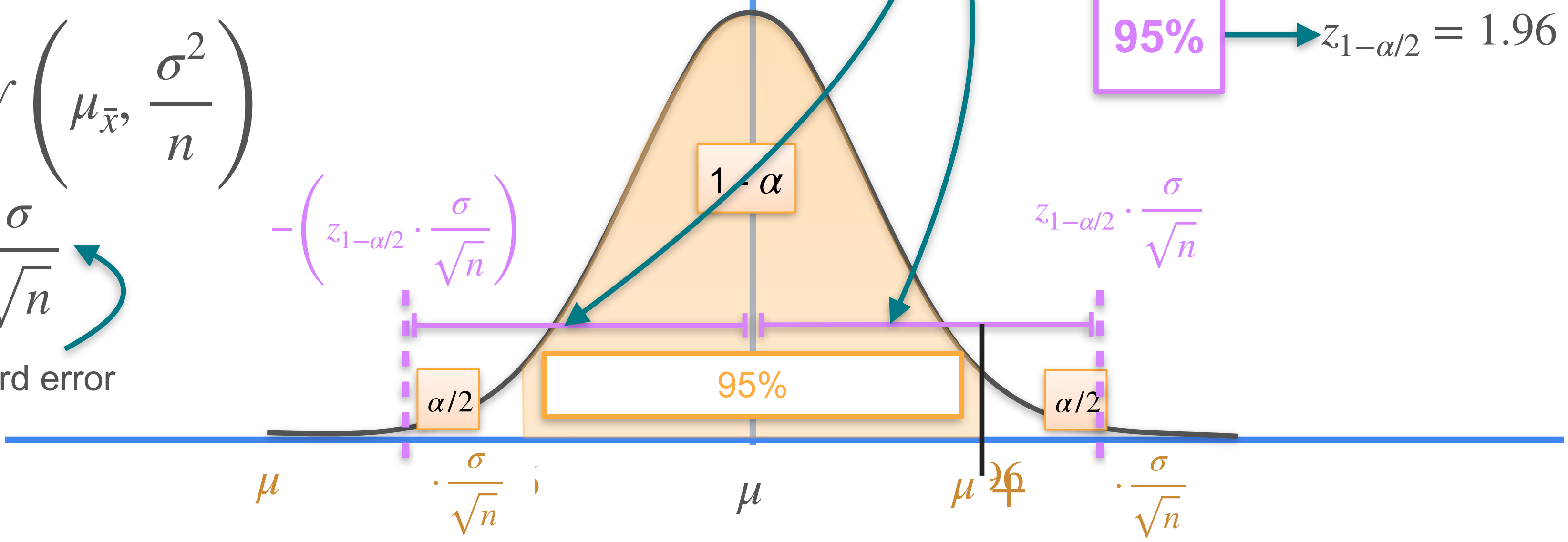
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Standard error

Critical Value

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

95% → $z_{1-\alpha/2} = 1.96$



DeepLearning.AI

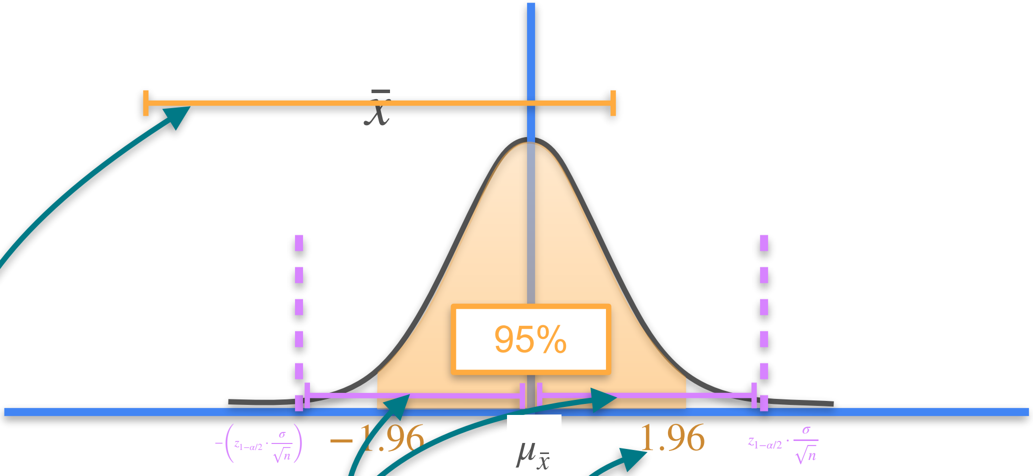
Confidence Interval

Confidence Interval - Calculation Steps

Confidence Interval - Calculation Steps

STEPS:

- Find the sample mean
- Define a desired confidence level (1 - α)
- Get the critical value ($z_{1-\alpha/2}$)
- Find the standard error ($\frac{\sigma}{\sqrt{n}}$)
- Find the margin of error
- Add/subtract the margin of error to the sample mean



$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$$\text{confidence interval} = \bar{x} \pm z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

Confidence Interval - Calculation Steps

STEPS:

- Find the sample mean
- Define a desired confidence level (1 - α)
- Get the critical value ($z_{1-\alpha/2}$)
- Find the standard error ($\frac{\sigma}{\sqrt{n}}$)
- Find the margin of error
- Add/subtract the margin of error to the sample mean

$$\text{confidence interval} = \bar{x} \pm z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

Assumptions

- Simple random sample
- Sample size > 30 or population is approximately normal

DeepLearning.AI

Confidence Interval

Confidence Interval - Example

Confidence Interval - Example

Statistopia

6,000 adults

Random Selection

49

$\bar{x} = 170cm$

$\sigma = 25cm$

95% $\rightarrow z_{1-\alpha/2} = 1.96$

Calculate a 95% confidence interval for the average height of adults on Statistopia.

Confidence Interval - Example

Random Selection

49

$\sigma = 25cm$

$\boxed{95\%} \rightarrow z_{1-\alpha/2} = 1.96$

Confidence Interval

$170cm \pm$ margin of error

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$$= 1.96 \bullet \frac{25}{\sqrt{49}}$$

$$= 1.96 \bullet \frac{25}{7}$$

$$= 7$$

Confidence Interval - Example

Random Selection

2500

$\sigma = 10cm$

$\boxed{95\%} \rightarrow z_{1-\alpha/2} = 1.96$

Confidence Interval

$170cm \pm$ margin of error

margin of error $= 7$

Confidence Interval

$170cm - 7 = 163cm$

$170cm + 7 = 177cm$

$163cm < \mu < 177cm$

DeepLearning.AI

Confidence Interval

Calculating Sample Size

Calculating Sample Size

$\boxed{\text{6,000 adults}}$

$\boxed{95\%} \rightarrow z_{1-\alpha/2} = 1.96$

$\bar{x} = 170cm \qquad \sigma = 25cm$

Random Selection

49

Margin of error: 7cm

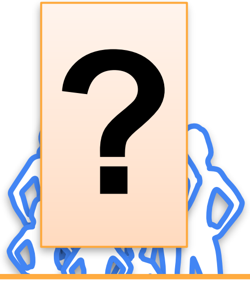
$\bar{x} \pm 7cm$

$163cm < \mu < 177cm$

Calculating Sample Size

6,000 adults

Random Selection



Margin of error: 7cm

$\bar{x} \pm 7cm$

$163cm < \mu < 177cm$

95% $\rightarrow z_{1-\alpha/2} = 1.96$

$\bar{x} = 170cm$ $\quad \sigma$

What is the smallest sample size to obtain the desired margin of error?

Margin of error: 3 cm

$\bar{x} \pm 3cm$

Calculating Sample Size

6,000 adults

95% $\rightarrow z_{1-\alpha/2} = 1.96$

$\bar{x} = 170cm \qquad \sigma = 25cm$

Margin of error: 3 cm

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$$3 = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$$3 \geq z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

Calculating Sample Size

6,000 adults

95% $\rightarrow z_{1-\alpha/2} = 1.96$

$\bar{x} = 170cm \qquad \sigma = 25cm$

Margin of error: 3 cm

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$$3 = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$$3 \geq z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

Calculating Sample Size

$\boxed{\text{6,000 adults}}$

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\boxed{95\%} \rightarrow z_{1-\alpha/2} = 1.96$

$$3 \geq z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\bar{x} = 170cm \qquad \sigma = 25cm$

$$3 \geq 1.96 \times \frac{25}{\sqrt{n}}$$

Margin of error: 3 cm

Calculating Sample Size

6,000 adults

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

95% $\rightarrow z_{1-\alpha/2} = 1.96$

$$3 \geq z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\bar{x} = 170cm \qquad \sigma = 25cm$

$$\frac{3}{1.96} \geq \frac{25}{\sqrt{n}}$$

Margin of error: 3 cm

Calculating Sample Size

$\boxed{\text{6,000 adults}}$

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\boxed{95\%} \rightarrow z_{1-\alpha/2} = 1.96$

$$3 \geq z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\bar{x} = 170cm \qquad \sigma = 25cm$

$$\frac{3}{1.96 \times 25} \geq \frac{1}{\sqrt{n}}$$

Margin of error: 3 cm

Calculating Sample Size

$\boxed{\text{6,000 adults}}$

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\boxed{95\%} \rightarrow z_{1-\alpha/2} = 1.96$

$$3 \geq z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\bar{x} = 170cm \qquad \sigma = 25cm$

$$\frac{1.96 \times 25}{3} \leq \frac{\sqrt{n}}{1}$$

Margin of error: 3 cm

Calculating Sample Size

6,000 adults

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

95% → $z_{1-\alpha/2} = 1.96$

$$\frac{1.96 \times 25}{3} \leq \frac{\sqrt{n}}{1}$$

$\bar{x} = 170cm \qquad \sigma = 25cm$

$$\left(\frac{1.96 \times 25}{3}\right)^2 \leq n$$

Margin of error: 3 cm

Calculating Sample Size

$\boxed{\text{6,000 adults}}$

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$\boxed{95\%} \longrightarrow z_{1-\alpha/2} = 1.96$

$$\frac{1.96 \times 25}{3} \leq \frac{\sqrt{n}}{1}$$

$\bar{x} = 170cm \qquad \sigma = 25cm$

$$n \geq \left(\frac{1.96 \times 25}{3}\right)^2$$

Margin of error: 3 cm

Calculating Sample Size

$\boxed{\text{6,000 adults}}$

margin of error $= z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$

$\boxed{95\%} \rightarrow z_{1-\alpha/2} = 1.96$

$$n \geq \left(\frac{1.96 \times 25}{3}\right)^2$$

$\bar{x} = 170cm \qquad \sigma = 25cm$

$$n \geq 266.78 \approx 267$$

Margin of error: 3 cm

Calculating Sample Size

$$\text{margin of error} = z_{1-\alpha/2} \bullet \frac{\sigma}{\sqrt{n}}$$

$$n \geq \left(\frac{1.96 \times 25}{3}\right)^2$$

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{MOE}\right)^2$$

DeepLearning.AI

Confidence Interval

Difference Between Confidence and Probability

Difference Between Confidence and Probability

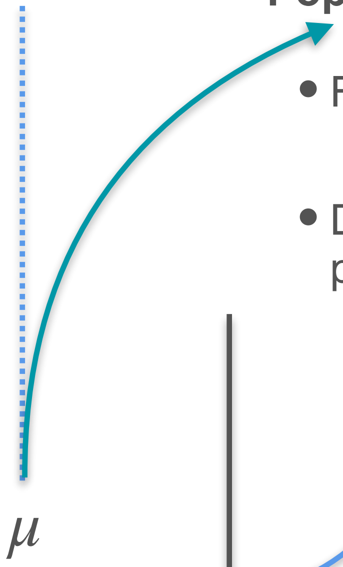
95% Confidence Level

The confidence interval contains the true population parameter approximately 95% of the time.

There's a 95% probability that the population parameter falls within the confidence interval.

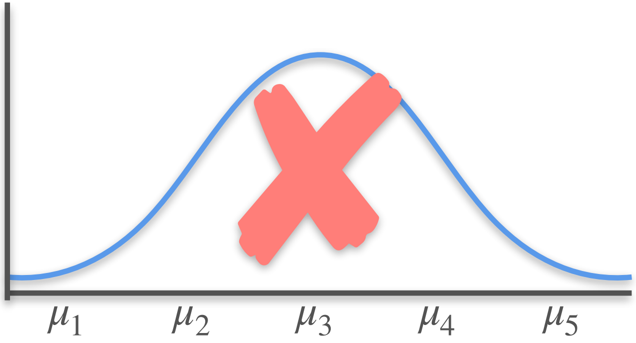
Difference Between Confidence and Probability

95% Confidence Level



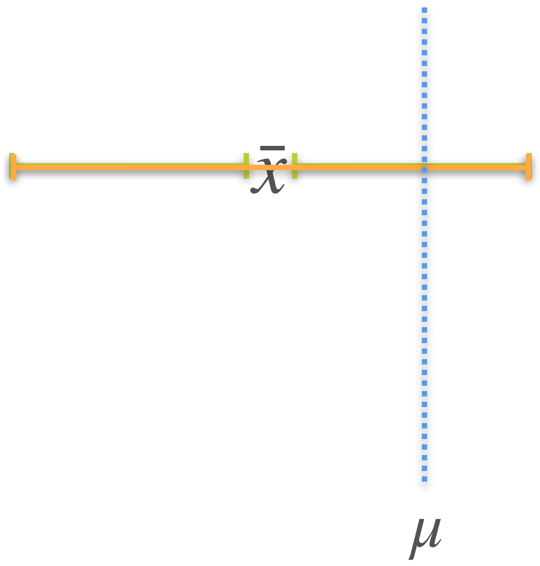
Population mean

- Fixed but unknown
- Does not have a probability distribution



Difference Between Confidence and Probability

95% Confidence Level

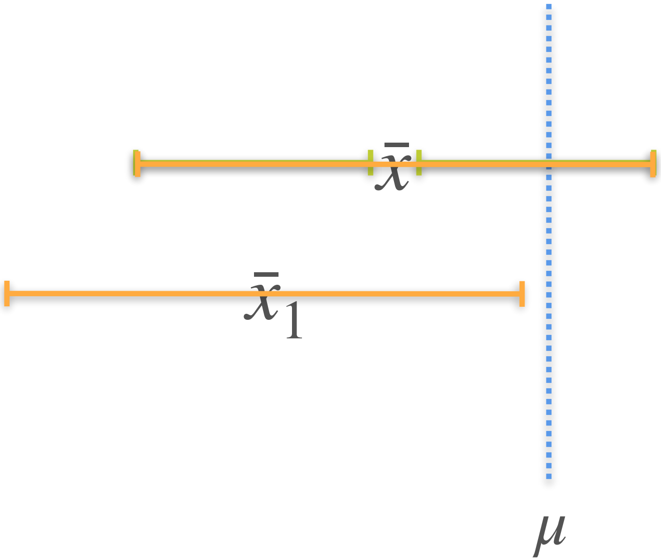


Population mean

- Fixed but unknown
- Does not have a probability distribution
- In the interval

Difference Between Confidence and Probability

95% Confidence Level

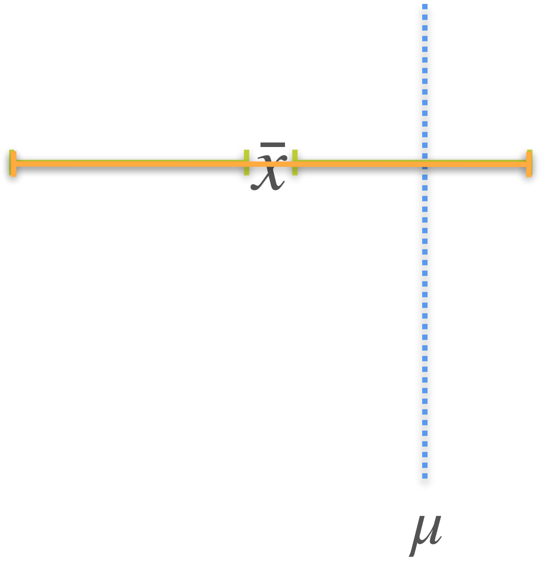


Population mean

- Fixed but unknown
- Does not have a probability distribution
- In the interval or not

Difference Between Confidence and Probability

95% Confidence Level



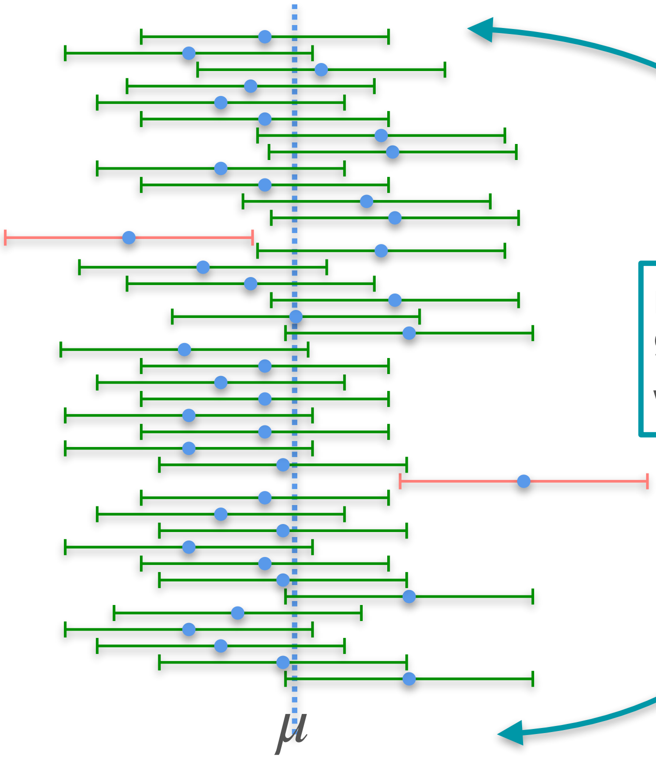
Population mean

- Fixed but unknown
- Does not have a probability distribution
- In the interval or not
- Does not fall within a specific interval 95% of the time

Difference Between Confidence and Probability

95% Confidence Level

success rate for constructing the confidence interval



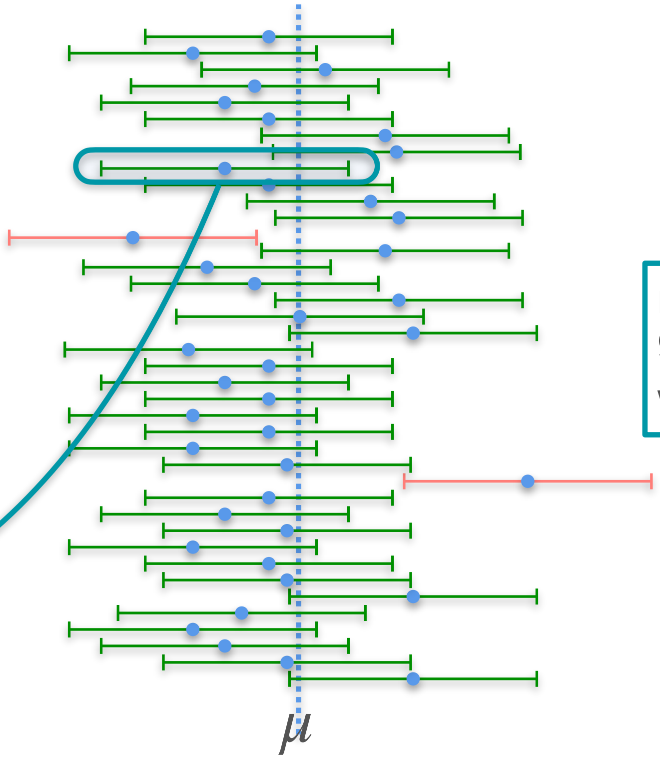
repeated many times, 95% of these intervals will contain μ

Difference Between Confidence and Probability

95% Confidence Level

success rate for constructing the confidence interval

not the probability that one specific intervals contains the population mean



repeated many times, 95% of these intervals will contain μ

DeepLearning.AI

Confidence Interval

Unknown Standard Deviation

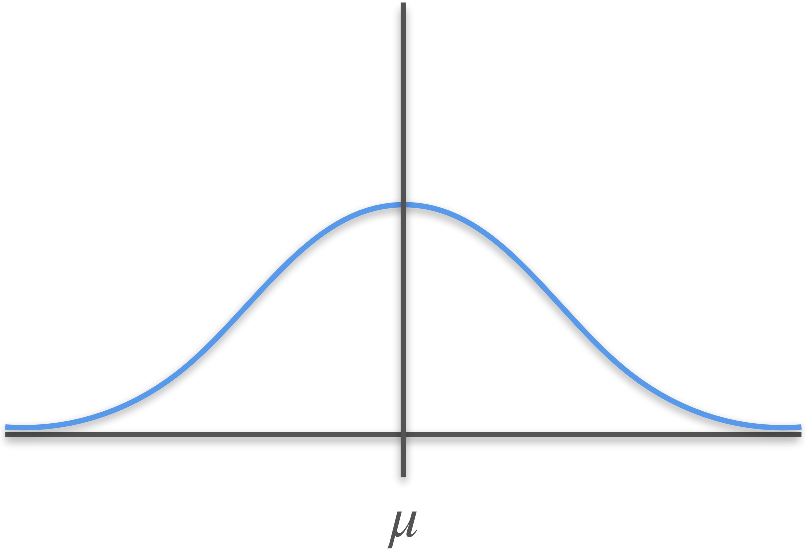
Confidence Interval - t Distribution

known σ

$$\bar{x} \quad \pm \quad \boxed{z_{1-\alpha/2}} \bullet \frac{\sigma}{\sqrt{n}}$$

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

normal distribution



Confidence Interval - t Distribution

known σ ? $\longrightarrow s$

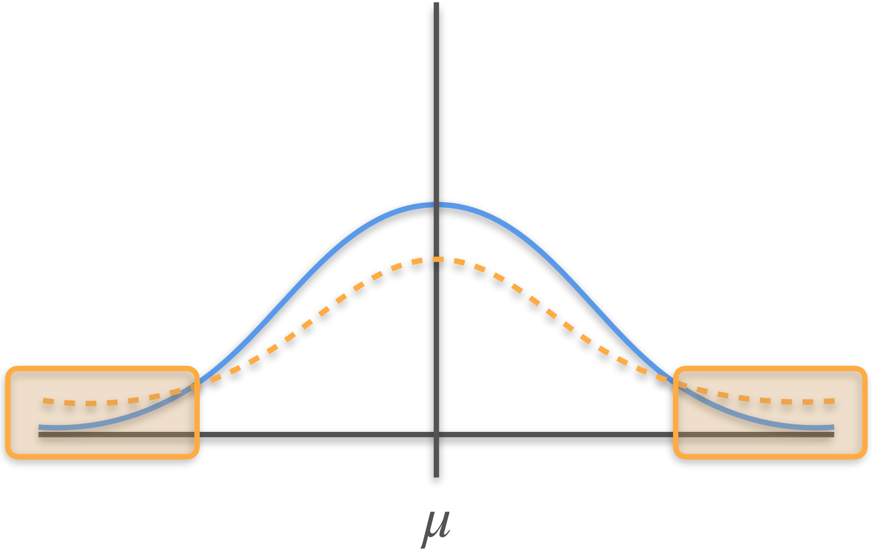
$$\bar{x} \quad \pm \quad \boxed{z_{1-\alpha/2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

normal distribution

not a normal distribution

student's t distribution



Confidence Interval - t Distribution

known σ

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

normal distribution

$$\bar{x} \pm \boxed{z_{1-\alpha/2}} \cdot \frac{\boxed{\sigma}}{\sqrt{n}}$$

unknown σ → replace with s

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

student's t distribution

$$\bar{x} \pm \boxed{z_{1-\alpha/2}} \cdot \frac{\boxed{s}}{\sqrt{n}}$$

Confidence Interval - t Distribution

known σ

$$\bar{x} \quad \pm \quad \boxed{z_{1-\alpha/2}} \cdot \frac{\boxed{\sigma}}{\sqrt{n}}$$

unknown σ

$$\bar{x} \quad \pm \quad \boxed{z_{1-\alpha/2}} \cdot \frac{\boxed{s}}{\sqrt{n}}$$

$$\bar{x} \quad \pm \quad \boxed{t_{1-\alpha/2}} \cdot \frac{s}{\sqrt{n}}$$

Confidence Interval - t Distribution

known σ

$$\bar{x} \pm \boxed{z_{1-\alpha/2}} \cdot \frac{\boxed{\sigma}}{\sqrt{n}}$$

unknown σ

$$\bar{x} \pm \boxed{t_{1-\alpha/2}} \cdot \frac{\boxed{s}}{\sqrt{n}}$$

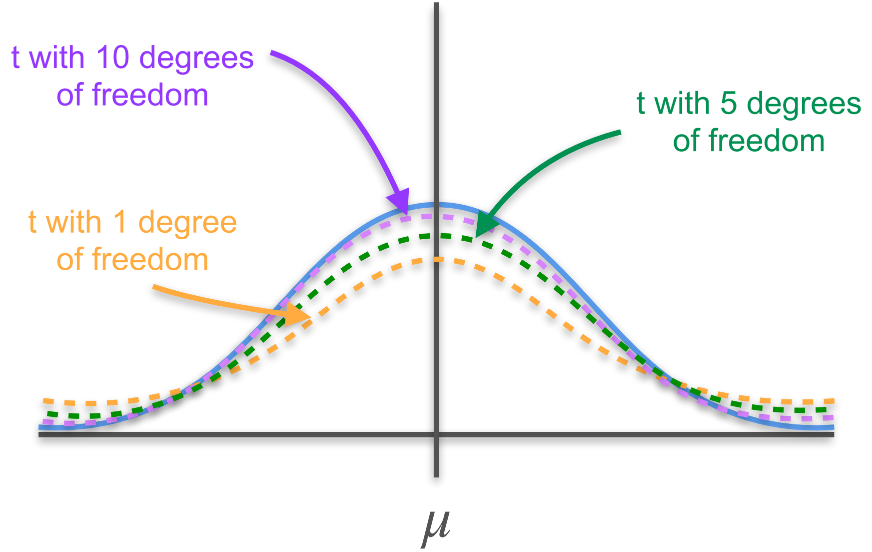
Confidence Interval - t Distribution

unknown σ

$$\bar{x} \pm \boxed{t_{1-\alpha/2}} \cdot \frac{\boxed{s}}{\sqrt{n}}$$

degrees of freedom

$$n-1$$



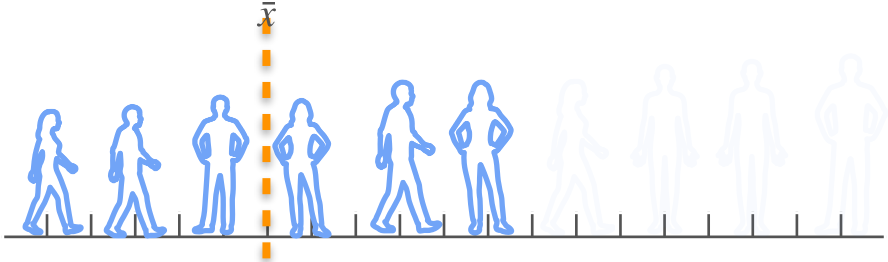
DeepLearning.AI

Confidence Interval

Confidence Intervals for Proportion

Confidence Interval for Proportions

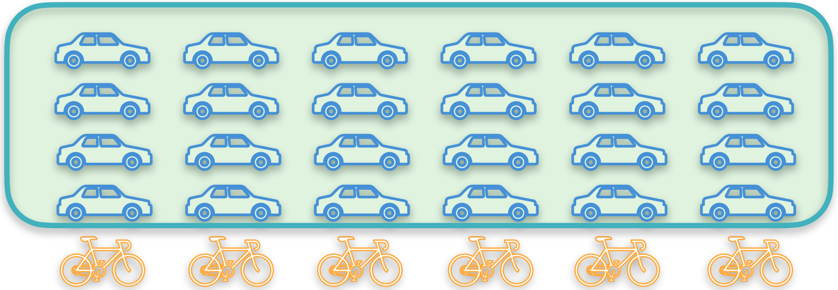
Confidence Interval for Means



confidence interval $= \bar{x} \pm$ margin of error

margin of error $= z_{1-\alpha/2} \bullet \dfrac{\sigma}{\sqrt{n}}$

Confidence Interval for Proportions



$x = 24$

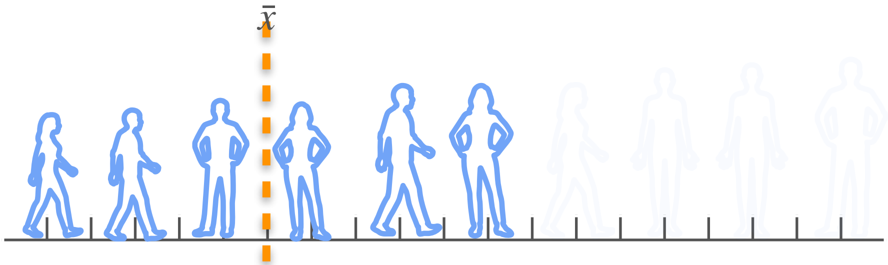
$n = 30$

$$\hat{p} = \frac{x}{n} = \frac{24}{30} = 80\,\%$$

How do you calculate a 95% confidence interval for this sample proportion?

Confidence Interval for Proportions

Confidence Interval for Means



confidence interval $= \bar{x} \pm$ margin of error

margin of error $= \boxed{z_{1-\alpha/2}} \cdot \boxed{\dfrac{\sigma}{\sqrt{n}}}$

Confidence Interval for Proportions

$$\hat{p} = \frac{x}{n} = \frac{24}{30} = 80\,\%$$

confidence interval $= \hat{p} \pm$ margin of error

margin of error $= \boxed{z_{1-\alpha/2}} \cdot \boxed{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}}$

standard error

Confidence Interval for Proportions

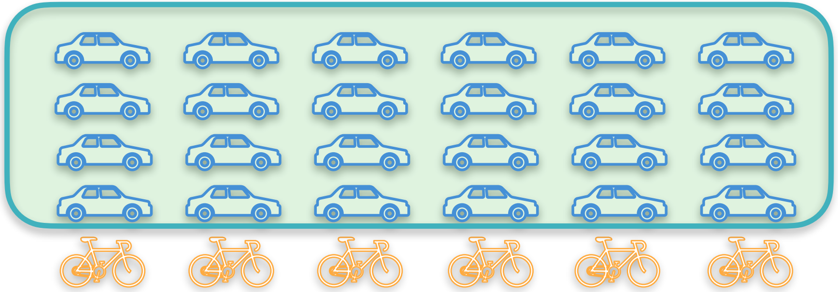
Confidence Interval for Proportions

confidence interval $= \hat{p} \pm$ margin of error

margin of error $= z_{1-\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

margin of error $= 1.96 \cdot \sqrt{\dfrac{0.8(1-0.8)}{30}}$

margin of error $= 0.14$



$n = 30$ $\quad \hat{p} = 80\,\% = 0.8$

Calculate a 95% confidence interval for this sample proportion

95% $\longrightarrow z_{1-\alpha/2} = 1.96$

Confidence Interval for Proportions

Confidence Interval for Proportions

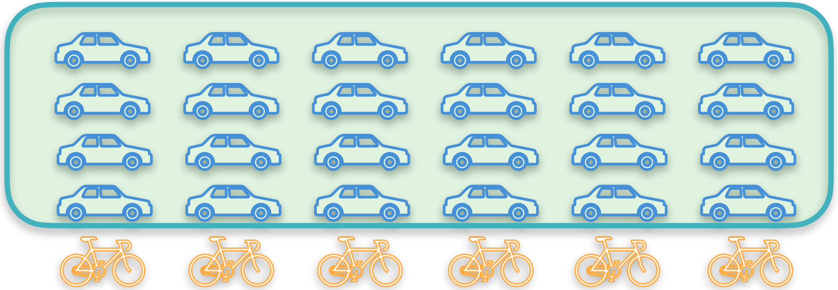
confidence interval $= \hat{p} \pm$ margin of error

margin of error $= 0.14$

confidence interval $= 0.8 \pm 0.14$

$$0.66 < p < 0.94$$

$$66\,\% < p < 94\,\%$$



$n = 30$ $\hat{p} = 80\,\% = 0.8$

Calculate a 95% confidence interval for this sample proportion

95% $\rightarrow z_{1-\alpha/2} = 1.96$

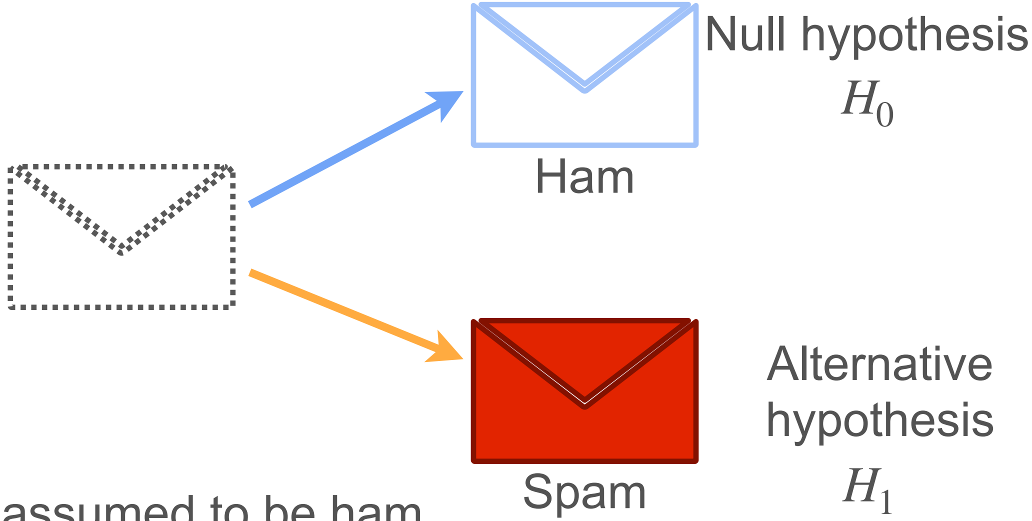
W4 Lesson 2

DeepLearning.AI

Hypothesis Testing

Defining Hypotheses

Motivation

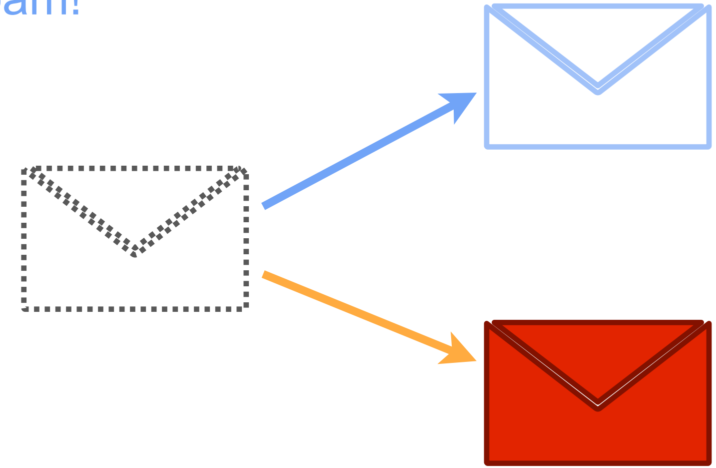


Email is assumed to be ham

Motivation

When rejecting that the email is not spam, you are accepting that the email is spam!

By failing to reject that the email IS ham, you are **not accepting** that it is actually ham



Null hypothesis

H_0

Mutually exclusive

True/False answer

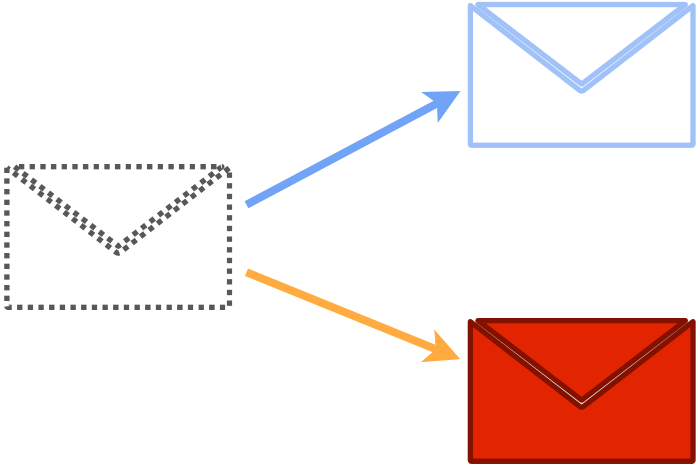
Alternative hypothesis

H_1

Email is assumed to be ham

Motivation

Not labeling the email spam, doesn't mean the email is ham!



Email is assumed to be ham

Null hypothesis

H_0

Baseline

Alternative hypothesis

H_1

What you want to prove

Data

Evidence

Sender

Attachments

Size

Keywords

How To Determine the Result of the Test

Plenty of evidence against H_0 → Reject H_0 (and accept H_1)

"Earn extra cash"

"Risk free"

"Dear friend"

"Act immediately"

"Apply now"

"Winner"

Very unlikely under H_0 (regular email) → Reject H_0 and send email to spam

DeepLearning.AI

Hypothesis Testing

Type I and Type II errors

Sometimes Things Go Wrong…

What if I make the wrong decision?

Type I error
(False positive)

Type II error
(False negative)

Type I and Type II Errors

Decision	Reality	
	H_0 True (Innocent)	H_0 False (Guilty)
Reject H_0 (Decide Guilty)	Type I error	Correct
Don't reject H_0 (Decide not guilty)	Correct	Type II error

Significance Level

Sending a regular email to spam is worse than sending a spam email to the regular inbox.

Type I error

Type II error

What is the greatest probability of type I error you are willing to tolerate?

Significance Level

What is the greatest probability of type I error you are willing to tolerate?

Significance level (α)

Type I error

Ham email identified as spam

0 — 1 — α

All emails are identified as **not spam** → No Type I error

Significance Level

What is the greatest probability of type I error you are willing to tolerate?

Significance level (α)

Type I error

Ham email identified as spam

Every time you get a regular email you make a Type I error

All emails are identified as **spam**

Significance Level

Significance level (α)

$\alpha = 0.05$

All emails are identified as **ham**

No Type I error

Both options are awful decision makers

You want a small α, but not zero

Catch:

↓Type I error ↑Type II error

All emails are identified as **spam**

Every time you get a regular email you make a Type I error

Significance Level

Type I error

Ham email identified as spam

$$\alpha = \max \mathbf{P}(\text{Type I error})$$
$$= \max \mathbf{P}(\text{Reject } H_0 \,|\, H_0)$$

The value of α is your criteria for designing your test

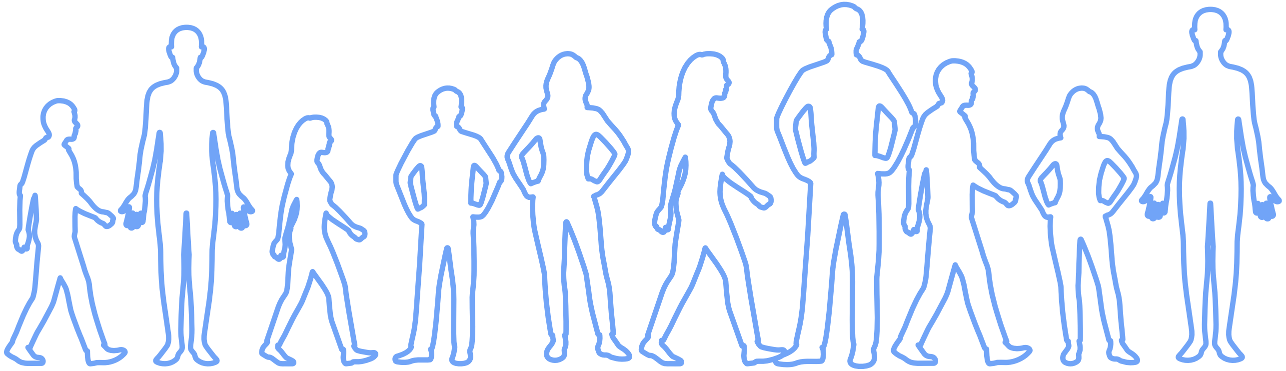
Given sample, α will determine if you reject H_0 or not

DeepLearning.AI

Hypothesis Testing

Right-Tailed, Left-Tailed and Two-Tailed tests

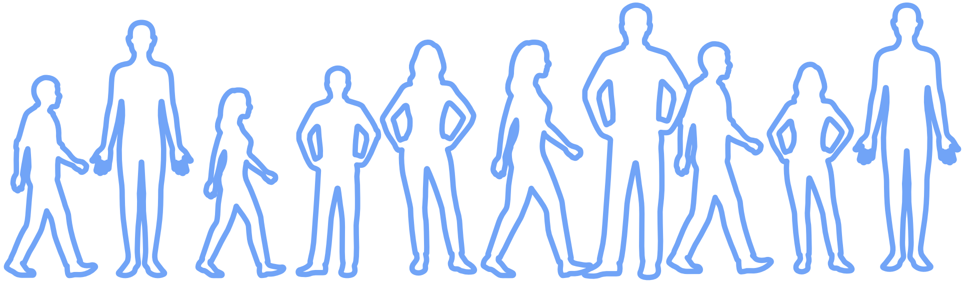
Example: Heights



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$\bar{x} = 68.442$$

Data Quality



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$\bar{x} = 68.442$$

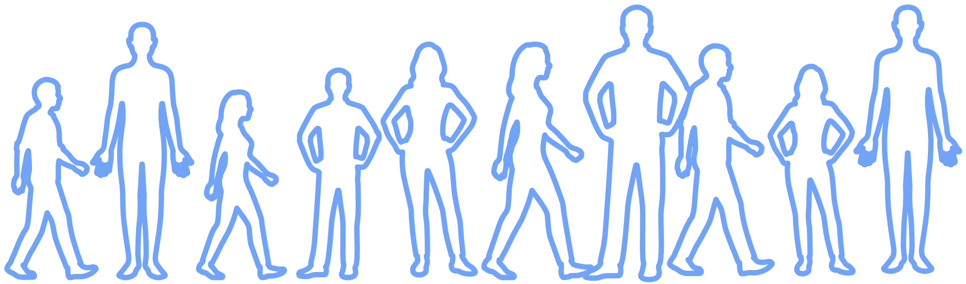
Reliable

Representative

Randomized

Sample size

Determining the Hypothesis



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

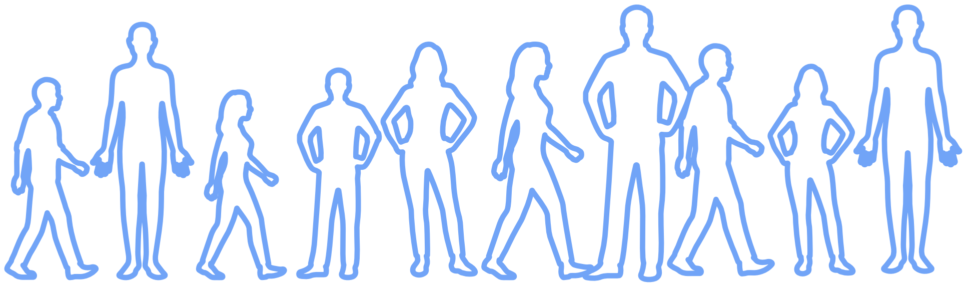
$\bar{x} = 68.442$

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Population

Test Statistic



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$\bar{x} = 68.442$

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

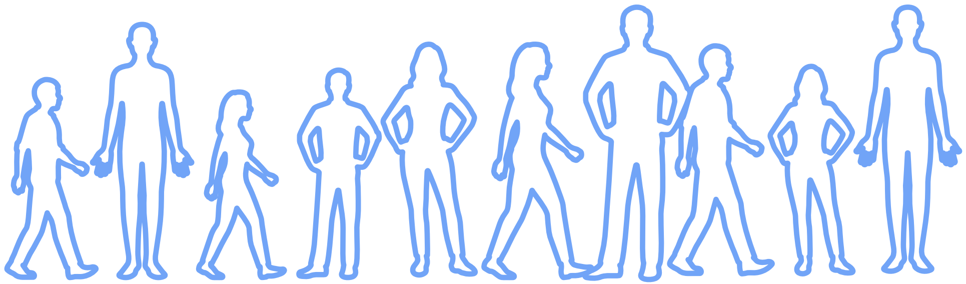
Observed statistic

Test statistic $\longrightarrow \bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$

Test Statistic

Test statistic

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$$



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$\bar{x} = 68.442$

Test statistic: $T(\boldsymbol{X})$ $\quad \boldsymbol{X} = (X_1, \ldots, X_n)$

Information about the population parameter under study

$\mu \rightarrow \overline{X}$ $\qquad p \rightarrow \overline{X}$ $\qquad \sigma^2 \rightarrow S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Not unique!

Example: Heights

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

3 questions

3 sets of hypothesis

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

$H_0 : \mu = 66.7$ vs. $H_1 : \mu < 66.7$

$H_0 : \mu = 66.7$ vs. $H_1 : \mu \neq 66.7$

Example: Heights

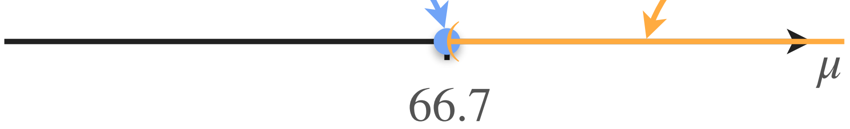
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\overline{X}$ Test statistic

Right-tailed test $H_0 : \mu \leq 66.7$ vs. $H_1 : \mu > 66.7$

Null hypothesis

Alternative hypothesis



If $\overline{x} \gg 66.7 \Rightarrow$ Reject H_0

Type I error: Determine $\mu > 66.7$, when population mean did not change

Type II error: Do not reject that $\mu = 66.7$ when in true $\mu > 66.7$

Example: Heights

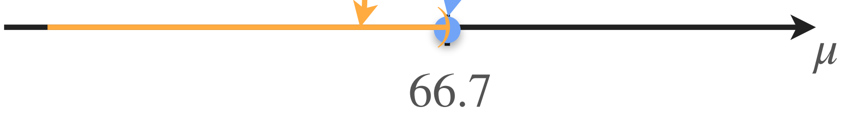
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\overline{X}$ Test statistic

Left tailed test $\longrightarrow$ $H_0 : \mu \geq 66.7$ vs. $H_1 : \mu < 66.7$

Alternative hypothesis

Null hypothesis



Type I error: Determine $\mu < 66.7$, when population mean did not change

If $\bar{x} \ll 66.7 \Rightarrow$ Reject H_0

Type II error: Don't reject that $\mu = 66.7$ when true $\mu < 66.7$

Example: Heights

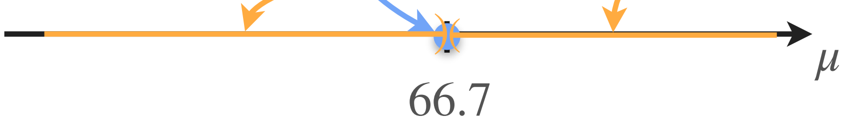
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\overline{X}$ Test statistic

Two tailed test $H_0 : \mu = 66.7$ vs. $H_1 : \mu \neq 66.7$

Null hypothesis

Alternative hypothesis



If $\bar{x} \gg 66.7$ or $\bar{x} \ll 66.7$ $\Rightarrow$ Reject H_0

Type I error: Determine $\mu \neq 66.7$, when population mean did not change

Type II error: Don't reject that $\mu = 66.7$ when true $\mu \neq 66.7$

Hypothesis Testing

p-Value

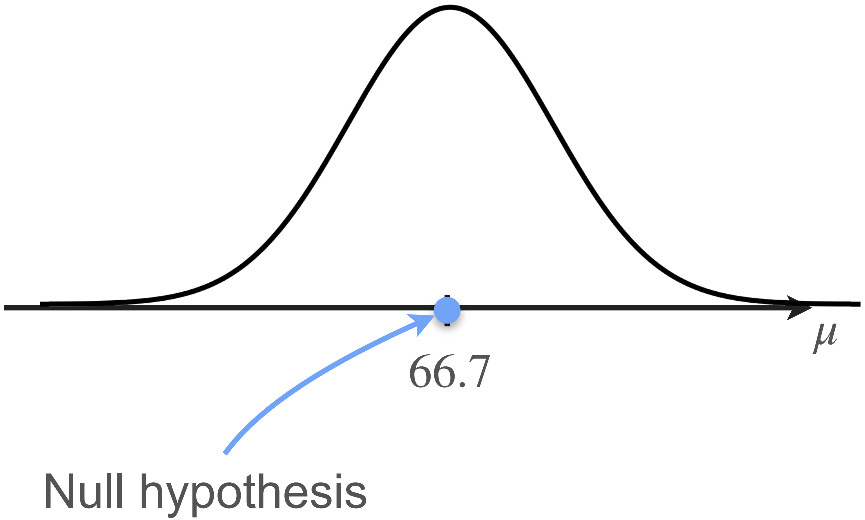
Example: Heights

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$

$n = 10$

If H_0 is true: $\overline{X} \sim \mathcal{N}\left(66.7, \frac{3^2}{10}\right)$



How likely was your sample if H_0 is true?

If the answer is very unlikely, then reject H_0

Right-Tailed Test for Gaussian Data (Known σ)

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$

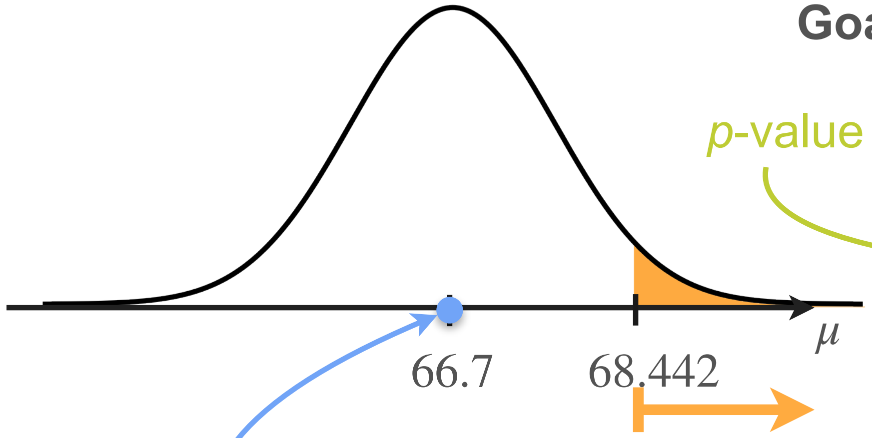
$n = 10$

$\bar{x} = 68.442$

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Goal: Type I error probability < $\alpha = 0.05$

Type I error: Determine $\mu > 66.7$, when population mean did not change



p-value

Null hypothesis

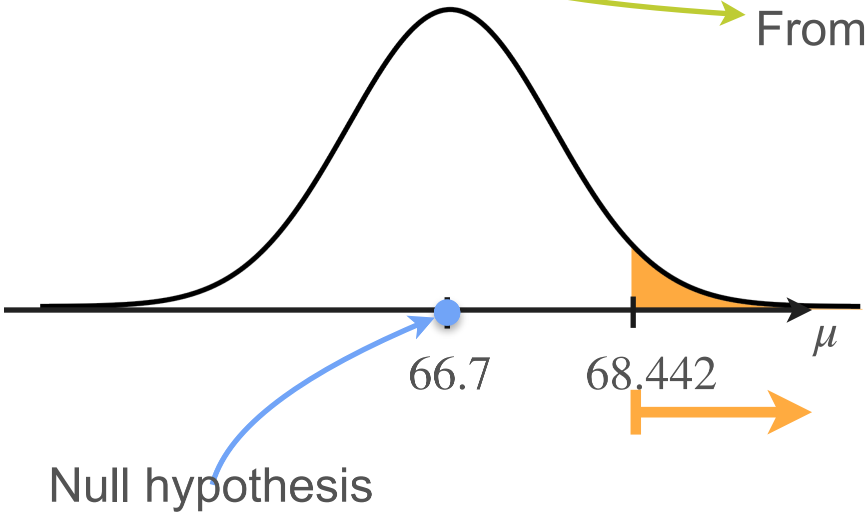
$$\mathbf{P}\left(\bar{X} > 68.442 \,\middle|\, \mu = 66.7\right) = 0.0332 < \alpha$$

Conclusion: reject H_0

(with a 5% significance level)

P-Values

A **p-value** is the probability, assuming H_0 is true, that the test statistic takes on a value as extreme as or more extreme than the value observed



From the observed value to the direction of H_1

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Decision rule:

If *p*-value $< \alpha$ reject H_0 (and accept H_1 as true)

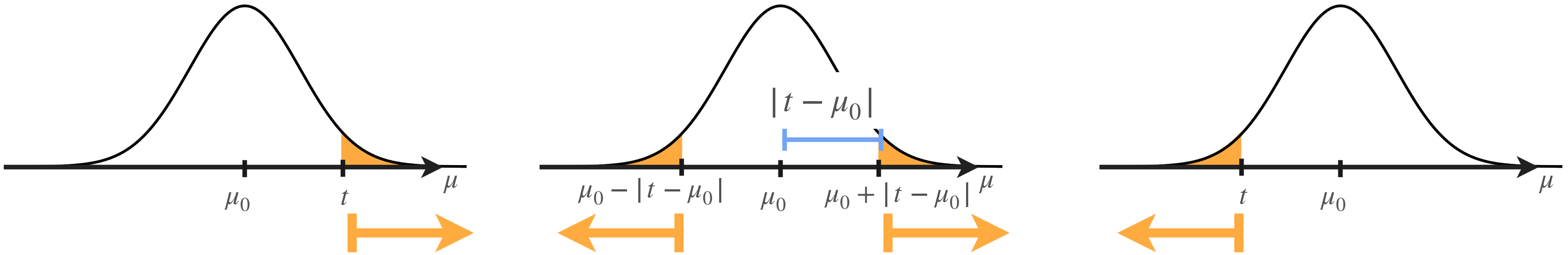
If *p*-value $> \alpha$ don't reject H_0

p-values

A **p-value** is the probability, assuming H_0 is true, that the test statistic takes on a value as extreme as or more extreme than the value observed

$T(X)$: test statistic $\quad$ t: observed statistic $\quad$ H_0: $\mu = \mu_0$

Right-tailed test $\quad$ Two-tailed test $\quad$ Left-tailed test



$\mathbf{P}(T(X) > t \,|\, H_0)$ $\quad$ $\mathbf{P}(|T(X) - \mu_0| > |t - \mu_0| \;|\, H_0)$ $\quad$ $\mathbf{P}(T(X) < t \,|\, H_0)$

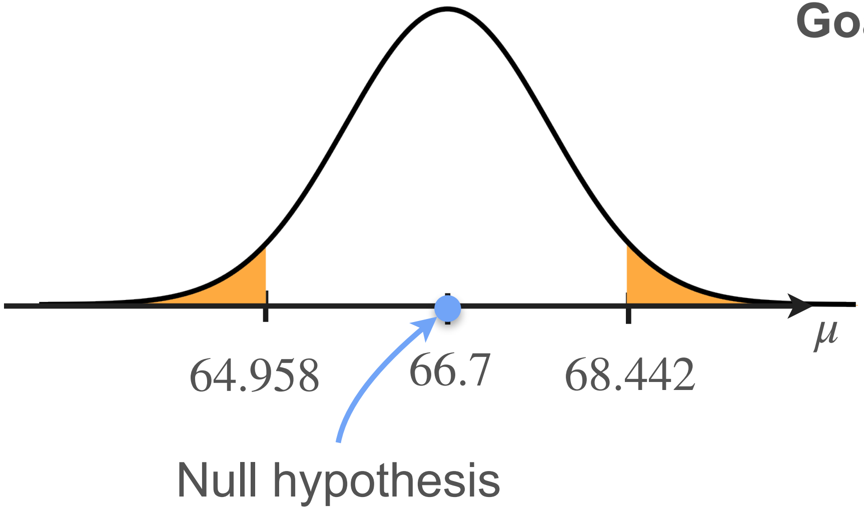
Two-Tailed Test for Gaussian Data (Known σ)

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$

$n = 10$

$\bar{x} = 68.442$



$H_0 : \mu = 66.7$ vs. $H_1 : \mu \neq 66.7$

Goal: Type I error probability < $\alpha = 0.05$

Type I error: Determine $\mu \neq 66.7$, when population mean did not change

$$\mathbf{P}\left(\left|\bar{X} - 66.7\right| > \left|68.442 - 66.7\right| \;\middle|\; \mu = 66.7\right)$$

$$= 0.0663 > \alpha$$

Conclusion: Do not reject H_0
(with a 5% significance level)

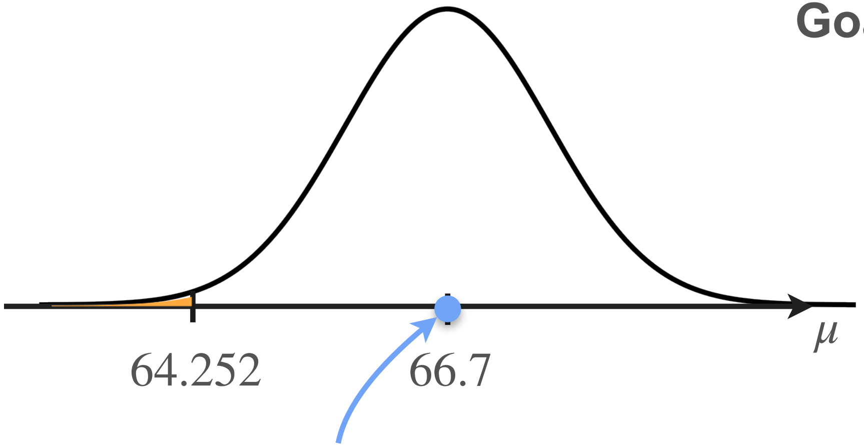
Left-Tailed Test for Gaussian Data (Known σ)

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$

$n = 10$

$\bar{x} = 64.252$



$H_0 : \mu = 66.7$ vs. $H_1 : \mu < 66.7$

Goal: Type I error probability < $\alpha = 0.05$

Type I error: Determine $\mu < 66.7$, when population mean did not change

$$\mathbf{P}\left(\bar{X} < 64.252 \,\middle|\, \mu = 66.7\right) = 0.0049 < \alpha$$

Conclusion: reject H_0
(with a 5% significance level)

Tests Using the Z-Statistic

So far, you used the statistic $\overline{X}$

$$\text{If } H_0 \text{ is true: } \overline{X} \sim \mathcal{N}\left(\mu_0, \frac{3^2}{10}\right)$$

Applying standardization, you can write equivalent tests using the

$$Z\text{-statistic } Z = \frac{\overline{X} - \mu_0}{3/\sqrt{10}}$$

$$\text{If } H_0 \text{ is true, } Z = \frac{\overline{X} - \mu_0}{3/\sqrt{10}} \sim \mathcal{N}(0,1)$$

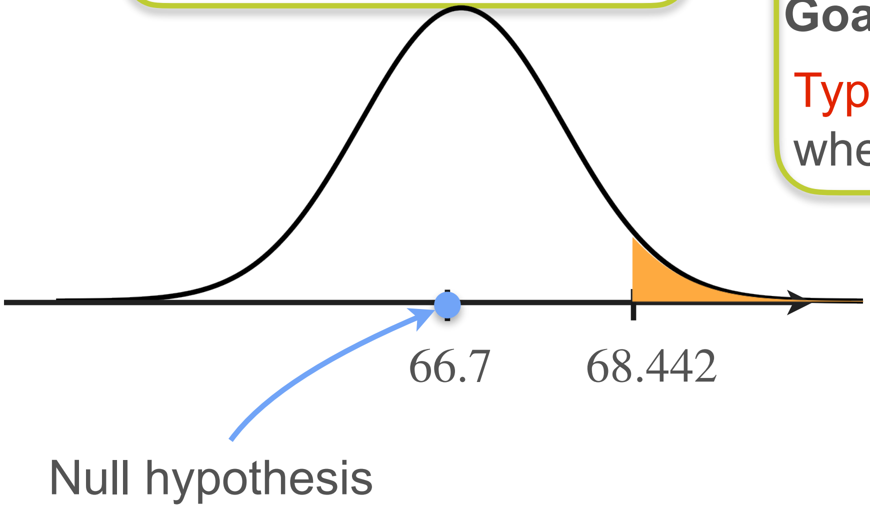
Right-Tailed Test Using the Z Statistic

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$

$n = 10$

$\bar{x} = 68.442$



$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Goal: Type I error probability < $\alpha = 0.05$

Type I error: Determine $\mu > 66.7$, when population mean did not change

$$\mathbf{P}\left(\bar{X} > 68.442 \,\middle|\, \mu = 66.7\right) = 0.0332 < \alpha$$

Conclusion: reject H_0

(with a 5% significance level)

Right-Tailed Test Using the Z Statistic

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$ $\quad \bar{x} = 68.442$

$n = 10$

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Goal: Type I error probability < $\alpha = 0.05$

Type I error: Determine $\mu > 66.7$, when population mean did not change

$$Z = \frac{\overline{X} - \mu_0}{3/\sqrt{10}} \longrightarrow z = \frac{68.442 - 66.7}{3/\sqrt{10}} = 1.837$$

$$\mathbf{P}\left(\overline{X} > 68.442 \,\middle|\, \mu = 66.7\right)$$

$$= 0.0332 < \alpha$$

$$\frac{\overline{X} - 66.7}{3/\sqrt{10}} > \frac{68.442 - 66.7}{3/\sqrt{10}} =$$

Conclusion: reject H_0

(with a 5% significance level)

Right-Tailed Test Using the Z Statistic

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$ $\quad \bar{x} = 68.442$

$n = 10$

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Goal: Type I error probability < $\alpha = 0.05$

Type I error: Determine $\mu > 66.7$, when population mean did not change

$$Z = \frac{\overline{X} - \mu_0}{3/\sqrt{10}} \rightarrow z = \frac{68.442 - 66.7}{3/\sqrt{10}} = 1.837$$

$$\frac{\overline{X} - 66.7}{3/\sqrt{10}} > \frac{68.442 - 66.7}{3/\sqrt{10}} = 1.837$$

$$\mathbf{P}\left(\quad > \quad \middle| \mu = 66.7 \right) = 0.0332 < \alpha$$

Conclusion: reject H_0

(with a 5% significance level)

Right-Tailed Test Using the Z Statistic

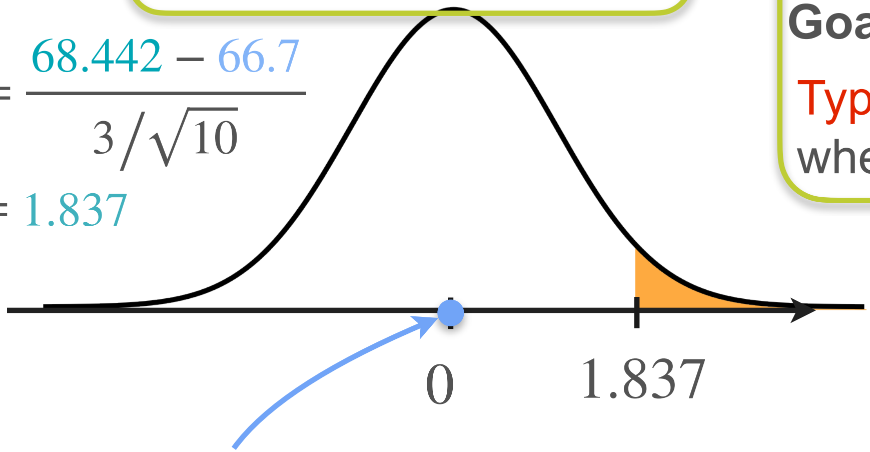
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$ $\quad \bar{x} = 68.442$

$n = 10$

$$z = \frac{68.442 - 66.7}{3/\sqrt{10}}$$

$$= 1.837$$



$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Goal: Type I error probability < $\alpha = 0.05$

Type I error: Determine $\mu > 66.7$, when population mean did not change

$$\mathbf{P}\left(\frac{\overline{X} - 66.7}{3/\sqrt{10}} > 1.837 \;\middle|\; \mu = 66.7 \right) = 0.0332 < \alpha$$

Conclusion: reject H_0
(with a 5% significance level)

DeepLearning.AI

Hypothesis Testing

Critical Values

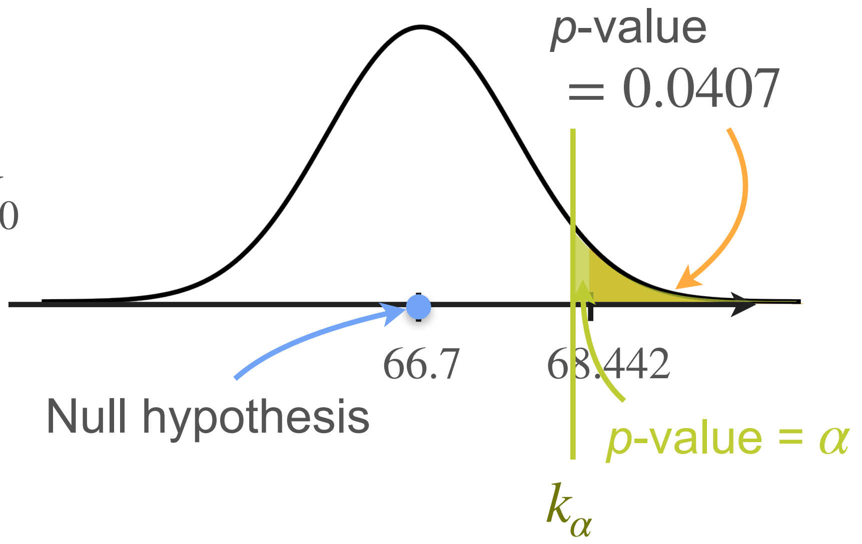
P-Values and Critical Values

If p-value $< \alpha$ ⇨ Reject H_0

What is the least extreme sample you could get that you would still reject H0?

Sample that has p-value = α

Critical values



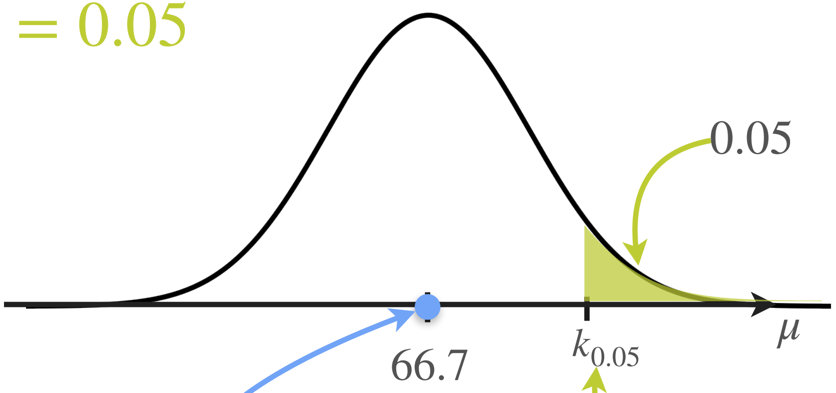
Computing Critical Values

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

Reject H_0

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

$\alpha = 0.05$



$n = 10 \qquad \sigma = 3 \qquad \bar{x} = 68.442$

$$0.05 = \mathbf{P}\left(\bar{X} > k_{0.05} \,\middle|\, \mu = 66.7\right)$$

If $\mu = 66.7$ $\bar{X} \sim \mathcal{N}\left(66.7, \frac{3^2}{10}\right)$

$k_{0.05} = 68.26$

Decision rule: Reject H_0 if $\bar{x} > 68.26$

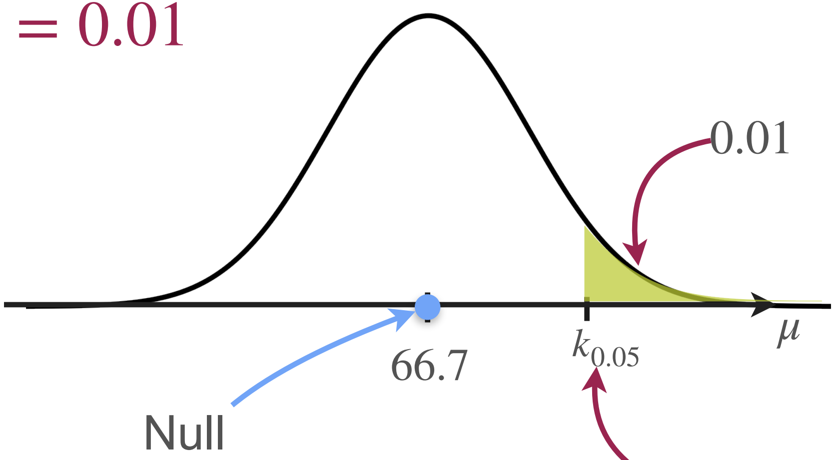
Computing Critical Values

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

Reject H_0

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

$\alpha = 0.01$



$n = 10 \qquad \sigma = 3 \quad \bar{x} = 68.442$

$$0.05 = \mathbf{P}\left(\bar{X} > k_{0.01} \,\middle|\, \mu = 66.7\right)$$

If $\mu = 66.7$ $\bar{X} \sim \mathcal{N}\left(66.7, \frac{3^2}{10}\right)$

$k_{0.05} = 68.26$

Decision rule: Reject H_0 if $\bar{x} > 68.26$

Computing Critical Values

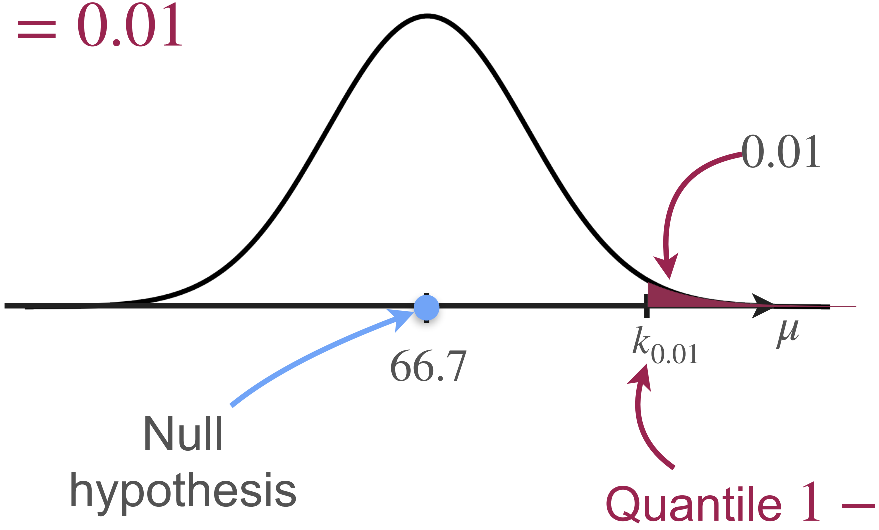
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

Do not reject H_0

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

$n = 10 \qquad \sigma = 3 \qquad \bar{x} = 68.442$

$\alpha = 0.01$



$$0.01 = \mathbf{P}\left(\bar{X} > k_{0.01} \,\middle|\, \mu = 66.7\right)$$

If $\mu = 66.7$ $\bar{X} \sim \mathcal{N}\left(66.7, \frac{3^2}{10}\right)$

$$k_{0.01} = 68.91$$

Decision rule: Reject H_0 if $\bar{x} > 68.91$

Critical Values

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$

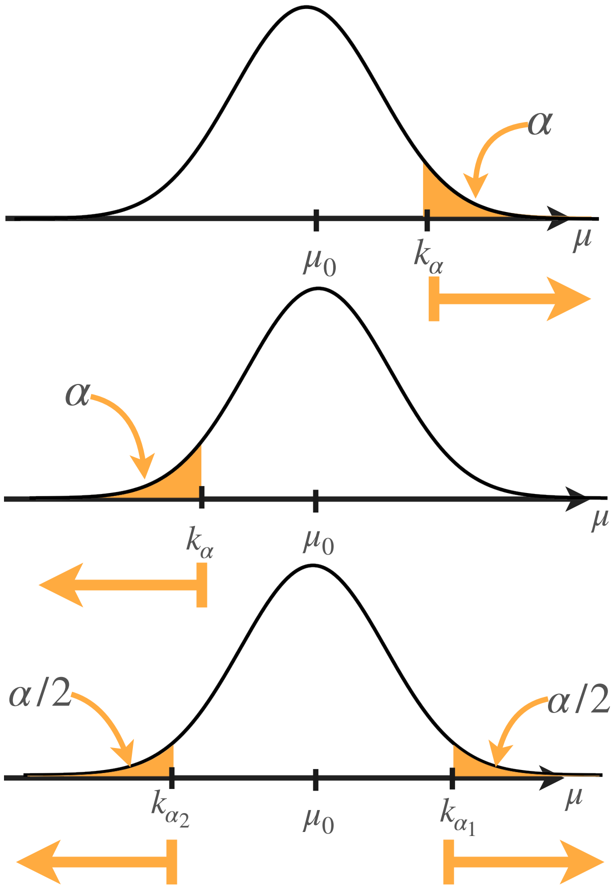
Decision rule: Reject H_0 if $t > k_\alpha$

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu < \mu_0$

Decision rule: Reject H_0 if $t < k_\alpha$

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

Decision rule: Reject H_0 if $t > k_{\alpha_1}$ or $t < k_{\alpha_2}$



k_α is the quantile $1 - \alpha$

k_α is the quantile α

k_{α_1} is the quantile $1 - \dfrac{\alpha}{2}$, and k_{α_2} is the quantile $\dfrac{\alpha}{2}$

Critical Values: Concluding Remarks

- You can define the critical value in advance
- For a given sample, using *p*-value and critical value will lead to the same conclusion
- Defining a test in terms of critical values makes determining Type II error probabilities for the decision rule.

DeepLearning.AI

Hypothesis Testing

Power of a Test

Type I and Type II Errors

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Decision	Reality	
	H_0 True ($\mu = 66.7$)	H_0 False ($\mu > 66.7$)
Reject H_0 (Decide $\mu > 66.7$)	Type I error	Correct
Don't reject H_0 (Decide $\mu = 66.7$)	Correct	Type II error

Finding the Type II Error Probabilities

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$ $\quad n = 10 \quad \sigma = 3$

For $\alpha = 0.05$: $k_\alpha = 68.26$ $\quad$ Decision rule: Reject H_0 if $\bar{x} > 68.26$

What is the **Type II error probability** if the true value is $\mu = 70$?

$\mathbf{P}\left(\text{Do not reject } H_0 \,|\, \mu = 70\right) \longrightarrow \mathbf{P}(\bar{X} < 68.26 \,|\, \mu = 70)$

Finding the Type II Error Probabilities

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$ $\quad n = 10 \quad \sigma = 3$

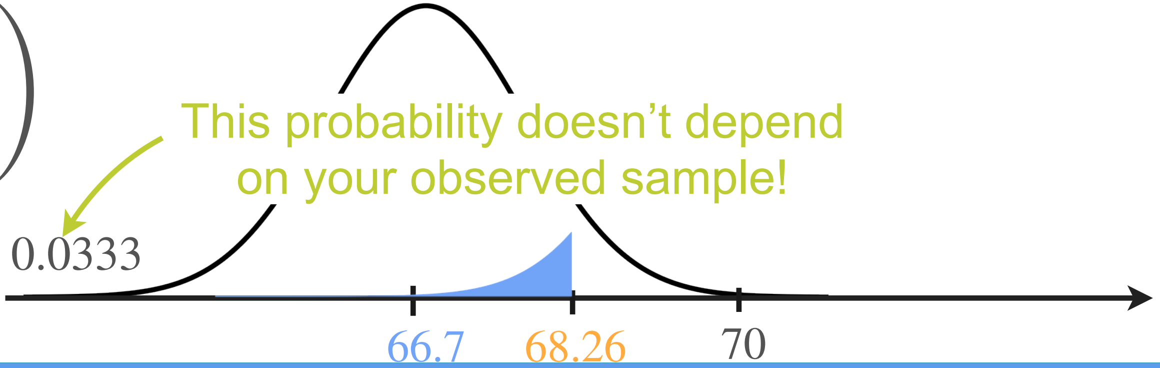
For $\alpha = 0.05$: $k_\alpha = 68.2604$ $\quad$ Decision rule: Reject H_0 if $\overline{X} > 68.26$

What is the **Type II error probability** if the true value is $\mu = 70$?

If $\mu = 70$ $\overline{X} \sim \mathcal{N}\left(70, \frac{3^2}{10}\right)$

$\beta = \mathbf{P}(\overline{X} < 68.26 \,|\, \mu = 70) = 0.0333$

This probability doesn't depend on your observed sample!



Power of the Test

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Decision	Reality	
	H_0 True ($\mu = 66.7$)	H_0 False ($\mu > 66.7$)
Reject H_0 (Decide $\mu > 66.7$)	Type I error	Correct
Don't reject H_0 (Decide $\mu = 66.7$	Correct	Type II error

Power of the test

$\mathbf{P}\left(\text{Reject } H_0 \,|\, \mu \in H_1\right)$

Power of the Test

$$\text{Type II error: } \overbrace{\mathbf{P}\left(\text{Do not reject } H_0 \,|\, \mu \in H_1\right)}^{\beta}$$

$$\text{Power of the test: } \underbrace{\mathbf{P}\left(\text{Reject } H_0 \,|\, \mu \in H_1\right)}_{1-\beta}$$

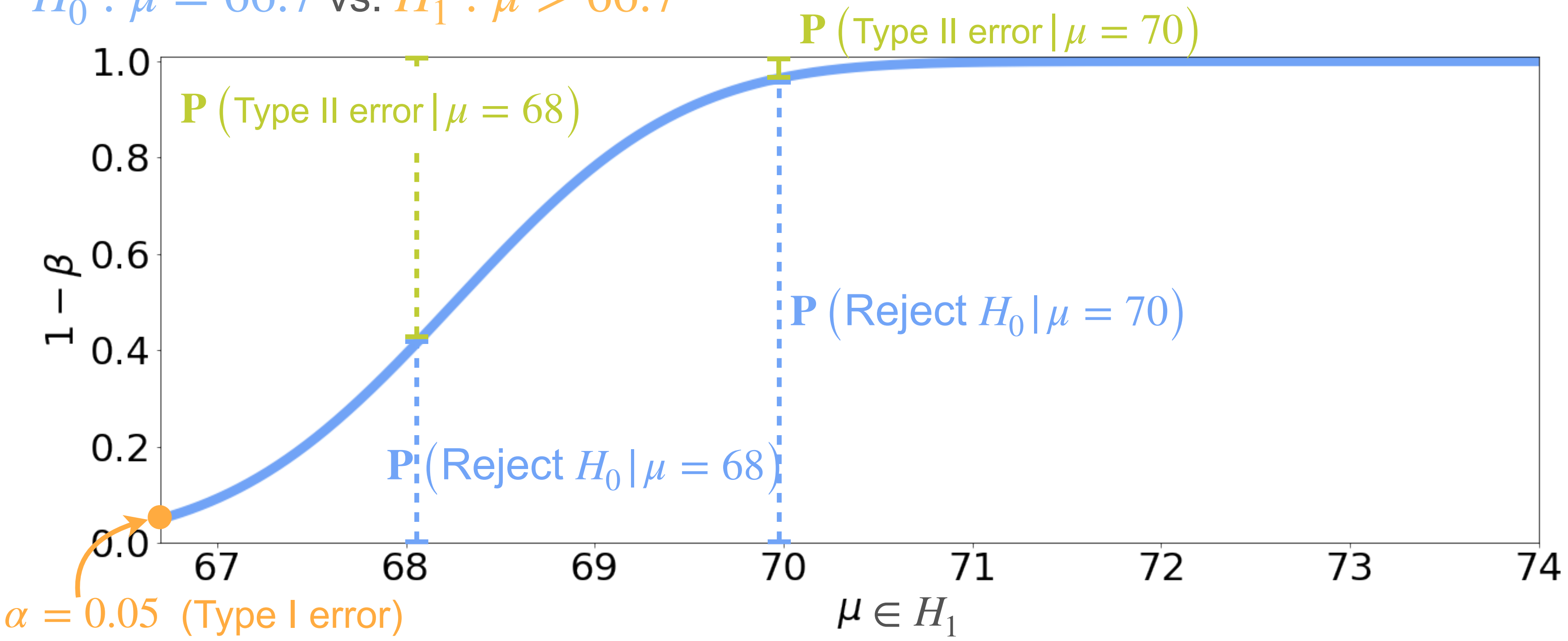
Complementary probabilities

$$\text{Power of the test } = 1 - \text{Type II error probability}$$

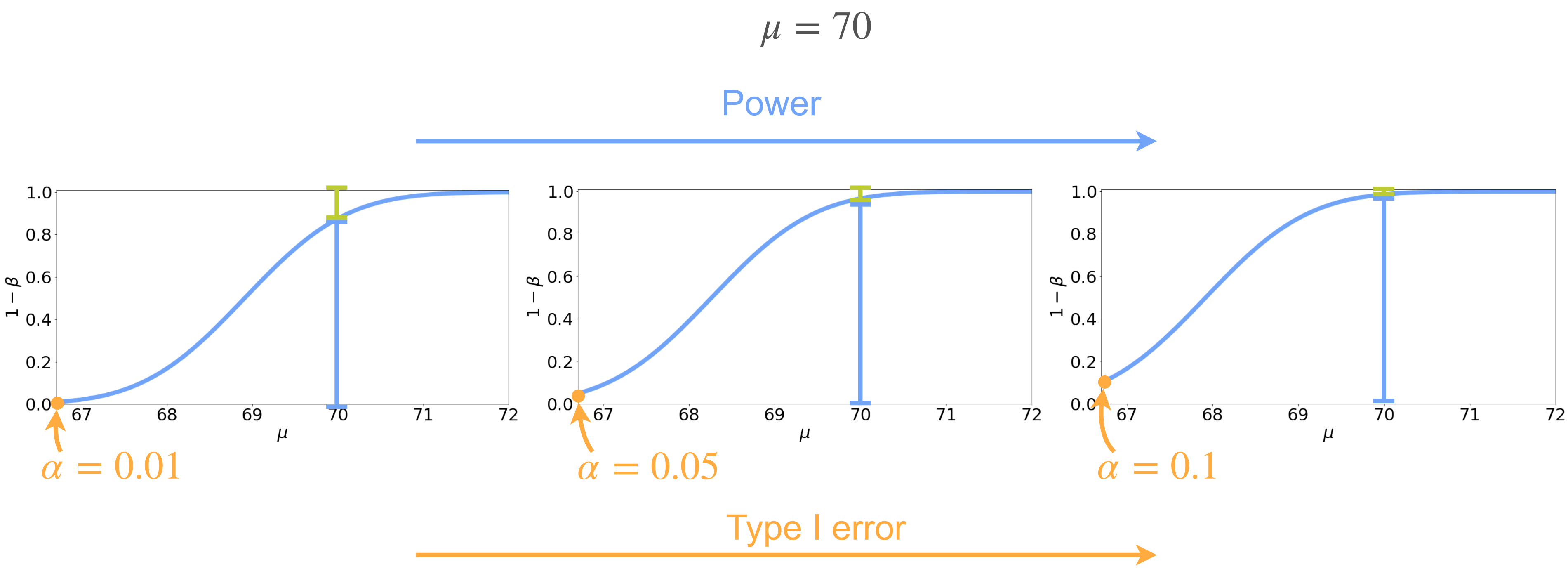
$$= 1 - \mathbf{P}\left(\text{Do not reject } H_0 \,|\, \mu \in H_1\right)$$

Power of the Test

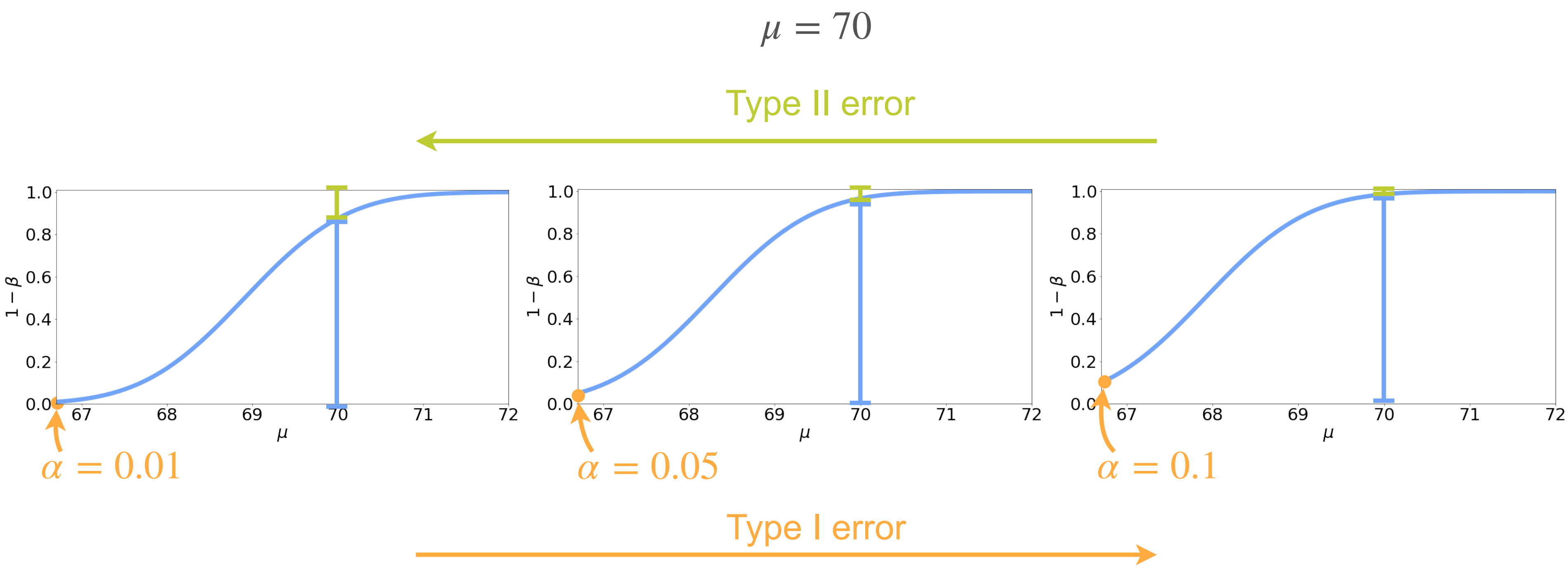
$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$



Power of the Test



Power of the Test



DeepLearning.AI

Hypothesis Testing

Interpreting results

Steps for Performing Hypothesis Testing

1. **State your hypotheses.**
    - Null hypothesis: the baseline $\longrightarrow H_0 : \mu = 66.7$
    - Alternative hypothesis: the statement you want to prove $\longrightarrow H_1 : \mu > 66.7$
2. **Design your test**
    - Decide the test statistic to work with. $\longrightarrow \bar{X}$
    - Decide the significance level $\longrightarrow \alpha = 0.05$
3. **Compute the observed statistic** (based on your sample) $\longrightarrow \bar{x} = 68.442$
4. **Reach a conclusion:**
    - If the *p*-value is less than the significance level reject H_0

$$\longrightarrow \mathbf{P}\left(\bar{X} > 68.442 \,|\, \mu = 66.7\right) \overset{?}{>} 0.05$$

Important Remarks - Interpreting Tests

- Type I error:
    - Reject null hypothesis when it as true → Reject H_0 when $\mu = 66.7$
- Type II error:
    - Do not reject null hypothesis, when if was false → Do not reject H_0 when $\mu > 66.7$
- Significance level (α):
    - It is the maximum probability of incurring in a Type I error
- Errors:
    - $\downarrow \alpha \uparrow \beta$

Important Remarks - Interpreting Tests

- *p*-values:
  - ~~If $\mathbf{P}(\text{Reject } H_0 \,|\, H_0) < \alpha \longrightarrow$ Reject H_0 and then accept H_1~~
  - ~~The *p*-value represents the probability of H_0 being true.~~

Important Remarks - Interpreting Tests

- *p*-values:
  - If $\mathbf{P}(\text{Reject } H_0 \,|\, H_0) < \alpha \longrightarrow$ Reject H_0 and then accept H_1
  - A small *p*-value indicates that the probability of seeing the observed data by chance is small

Important Remarks - Interpreting Tests

- *p*-values:
  - If $\mathbf{P}(\text{Reject } H_0 \,|\, H_0) < \alpha \longrightarrow$ Reject H_0 and then accept H_1
  - A small *p*-value indicates that the probability of seeing the observed data by chance is small
- Test conclusions
  - Reject $H_0 \rightarrow H_1$ true
  - Do not reject $H_0 \rightarrow$ ~~H_0 true~~

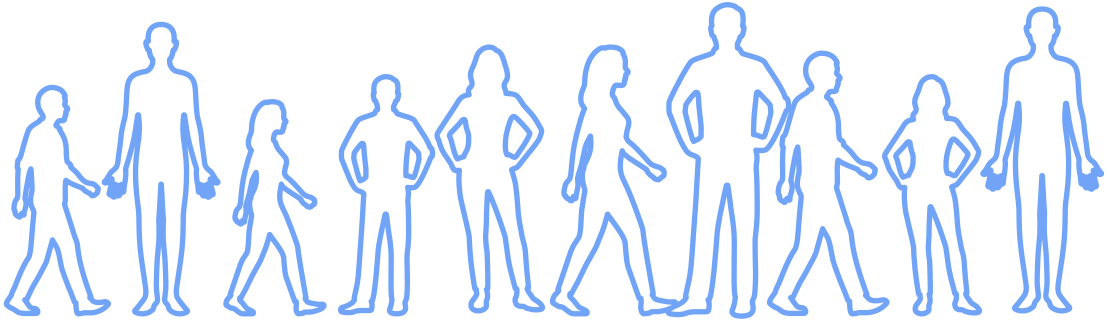
You can only say that there is not enough evidence

DeepLearning.AI

Hypothesis Testing

t-Distribution

t-Distribution: Motivation



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$X_i \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{10}\right)$$

This is fine if you know μ and σ

What if σ is unknown?

t-Distribution: Motivation

$$X_i \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{10}\right)$$

$$S = \sqrt{\frac{1}{10-1}\sum_{i=1}^{10}(X_i - \bar{X})^2}$$

If μ, σ are known

$\boxed{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{10}}} \sim \mathcal{N}(0,1^2)$ (Standardization)

Z statistic

T statistic

What if σ is unknown?

Replace σ with its estimate $\boxed{\dfrac{\bar{X}-\mu}{S/\sqrt{10}}} \sim \cancel{\mathcal{N}(0,1^2)}$

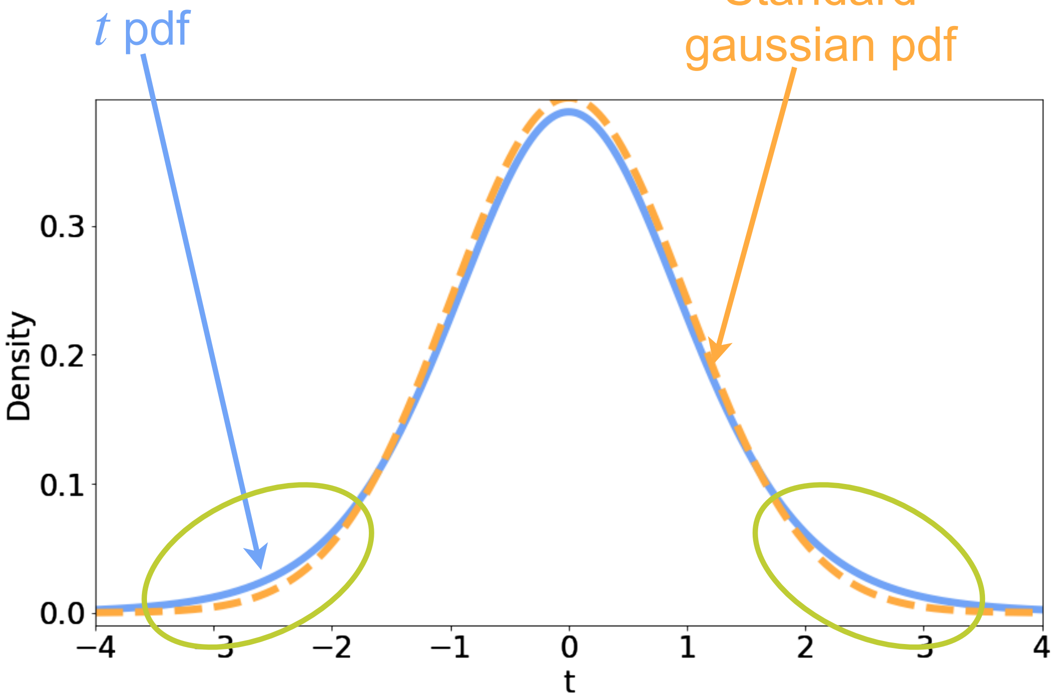
t-Distribution

$\dfrac{\overline{X}-\mu}{S/\sqrt{10}}$ follows a t distribution

What does it look like?

Still bell-shaped

It has heavier tails that account for the uncertainty introduced with the std estimation



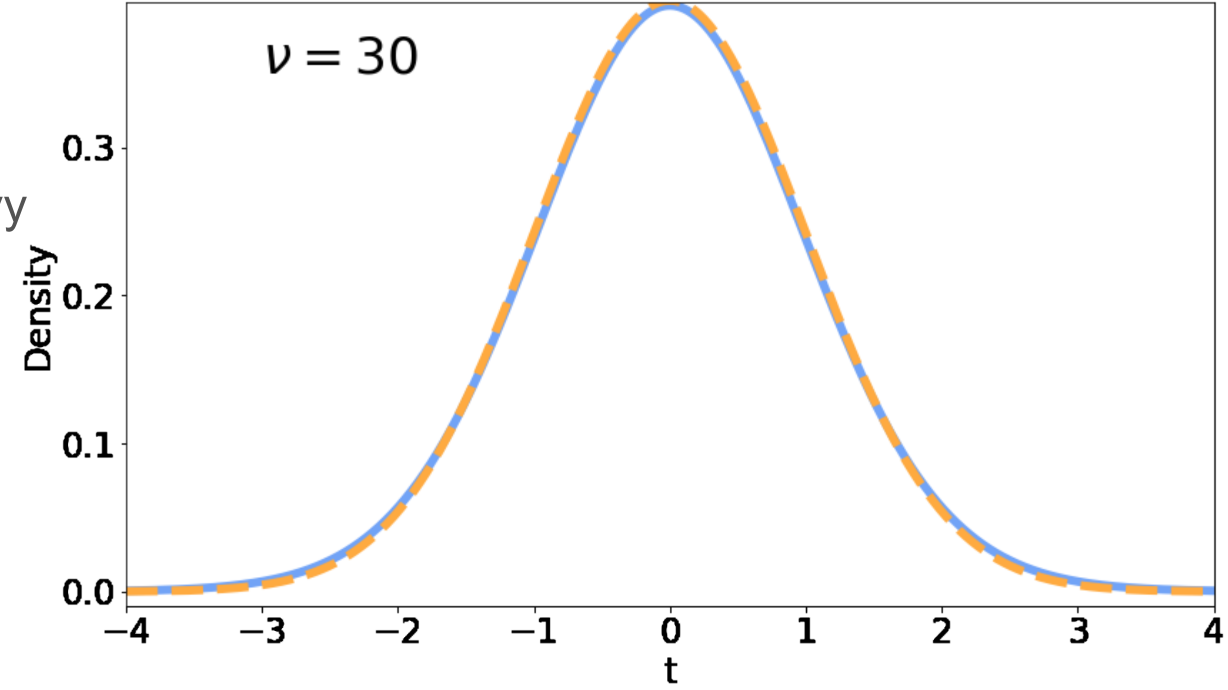
t-Distribution

Parameters:

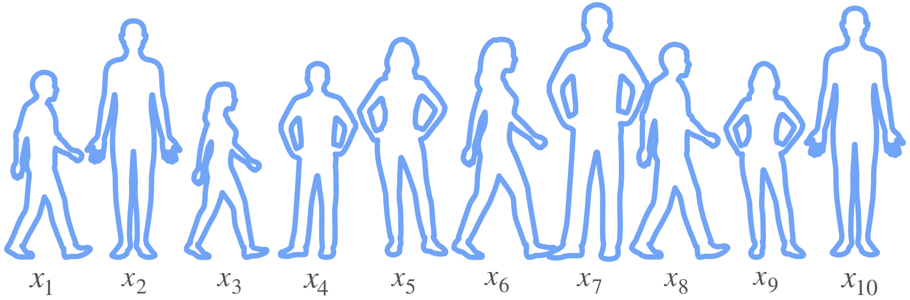
- Degrees of freedom (ν)

Controls how heavy the tails are

$X \sim t_\nu$



t-Distribution and T-Statistic



$n = 10$

$\nu = 10 - 1$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{10}} \sim t_9$$

Degrees of freedom (ν) = sample size -1
= $(n - 1)$

As n increases, this looks more like a $\mathcal{N}(0,1^2)$

T-statistic is used when

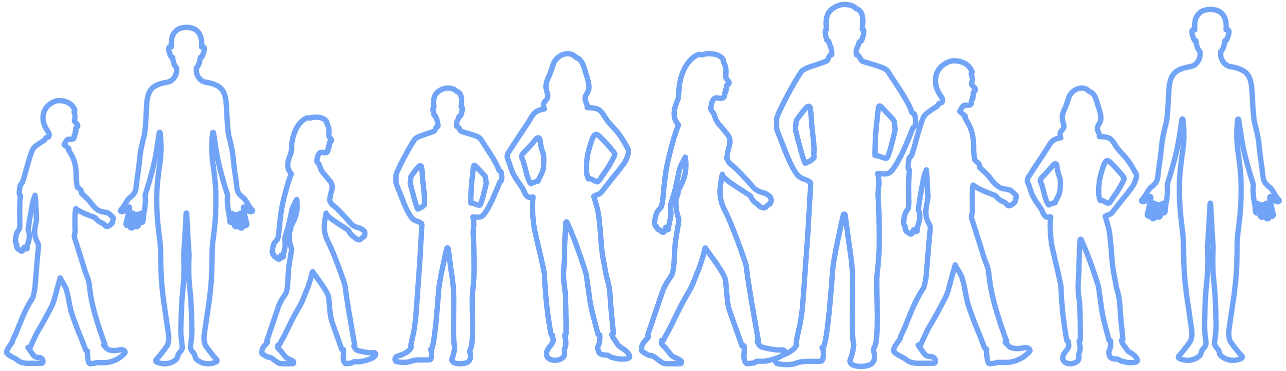
- The population has a Gaussian distribution
- But you don't know the variance

DeepLearning.AI

Hypothesis Testing

t-Tests

Example: Heights



x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

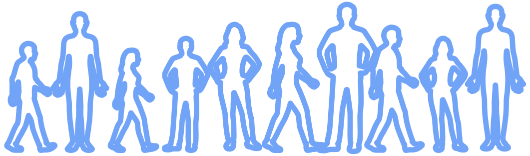
$$\bar{x} = 68.442$$

Example: Heights

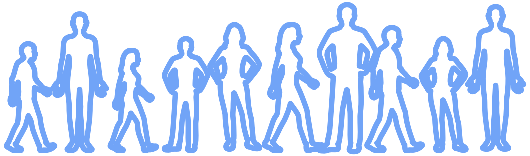
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

3 questions

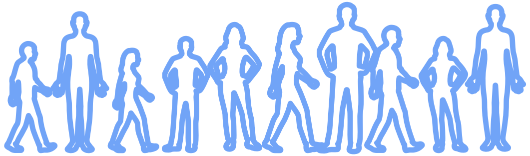
3 sets of hypothesis



$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$



$H_0 : \mu = 66.7$ vs. $H_1 : \mu < 66.7$



$H_0 : \mu = 66.7$ vs. $H_1 : \mu \neq 66.7$

Example: Heights

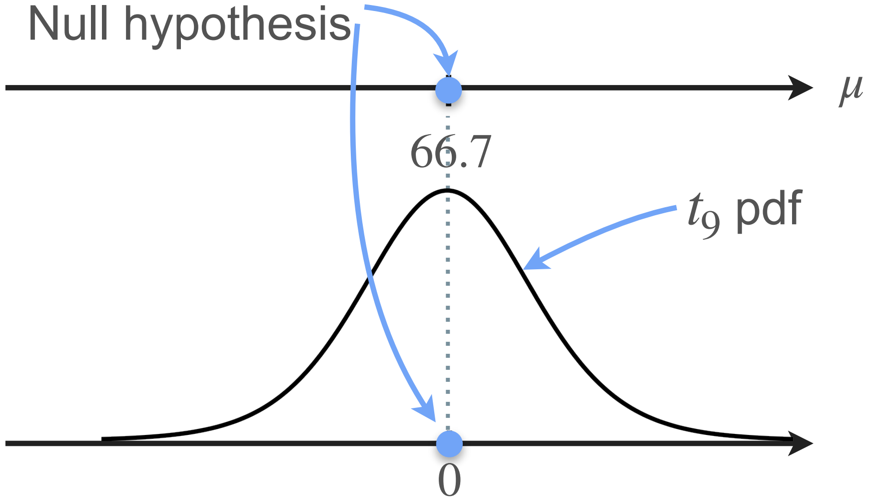
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\cancel{\sigma = 3}$

$n = 10$

$H_0 : \mu = 66.7$

$\cancel{\text{If } H_0 \text{ is true: } \bar{X} \sim}$



If H_0 is true: $T = \dfrac{\bar{X} - 66.7}{S/\sqrt{10}} \sim t_9$

Right-Tailed Test for Gaussian Data (Unknown σ)

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

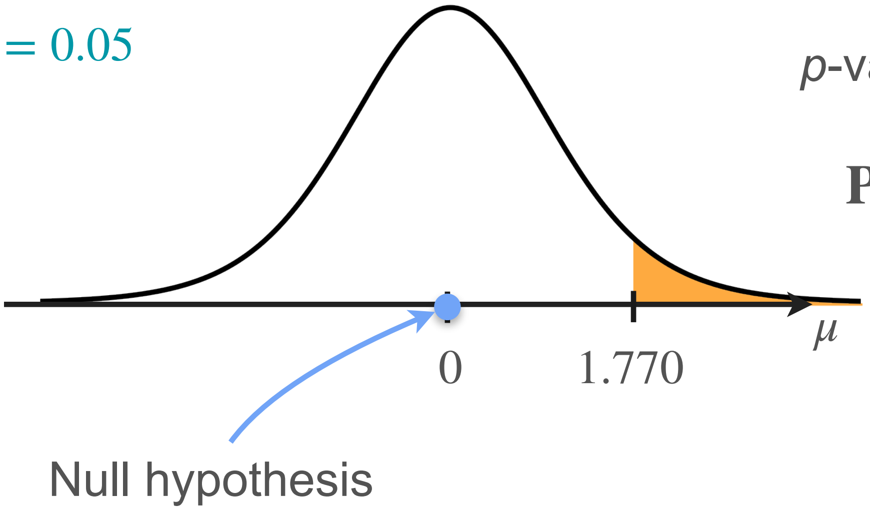
$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$ $\quad n = 10$

$\bar{x} = 68.442$

$s = 3.113$

$$t = \frac{68.442 - 66.7}{3.113/\sqrt{10}} = 1.770$$

$\alpha = 0.05$



p-value:

$$\mathbf{P}\left(\frac{\bar{X} - 66.7}{S/\sqrt{10}} > 1.770 \,\middle|\, \mu = 66.7\right)$$

$$= 0.0552 > \alpha$$

Conclusion: do not reject H_0

(with a 5% significance level)

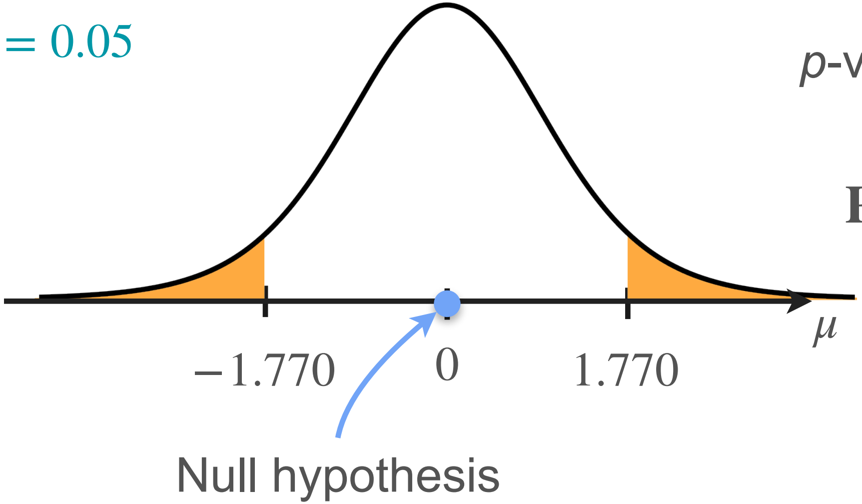
Two-Tailed Test for Gaussian Data (Unknown σ)

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$H_0 : \mu = 66.7$ vs. $H_1 : \mu \neq 66.7$ $\quad n = 10$ $\quad \bar{x} = 68.442$ $\quad s = 3.113$

$$t = \frac{68.442 - 66.7}{3.113/\sqrt{10}} = 1.770$$

$\alpha = 0.05$



p-value:

$$\mathbf{P}\left(\left|\frac{\bar{X} - 66.7}{S/\sqrt{10}}\right| > |1.770| \,\middle|\, \mu = 66.7\right) = 0.1105 > \alpha$$

Conclusion: do not reject H_0

(with a 5% significance level)

Left-Tailed Test for Gaussian Data (Unknown σ)

The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

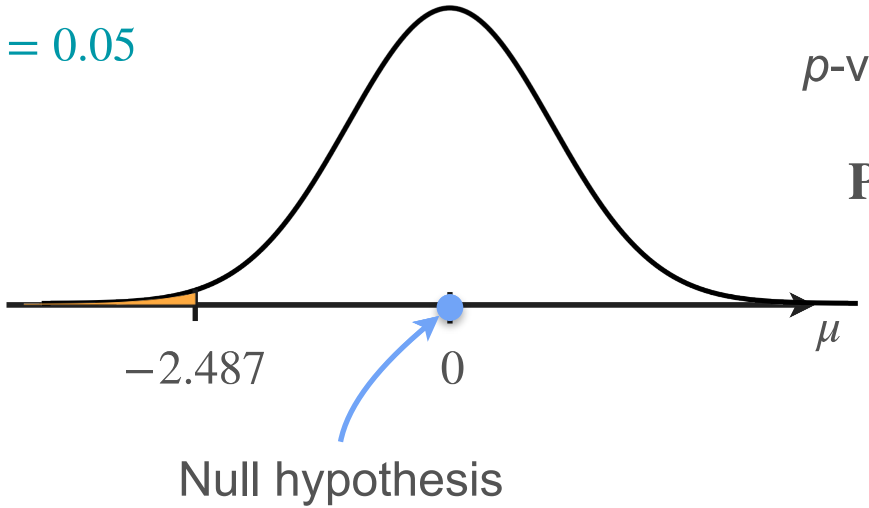
$H_0 : \mu = 66.7$ vs. $H_1 : \mu < 66.7$ $n = 10$

$\bar{x} = 64.252$

$s = 3.113$

$$t = \frac{64.252 - 66.7}{3.113/\sqrt{10}} = -2.487$$

$\alpha = 0.05$



p-value:

$$\mathbf{P}\left(\frac{\bar{X} - 66.7}{S/\sqrt{10}} < -2.487 \,\middle|\, \mu = 66.7\right)$$

$$= 0.0173 < \alpha$$

Conclusion: reject H_0
(with a 5% significance level)

DeepLearning.AI

Hypothesis Testing

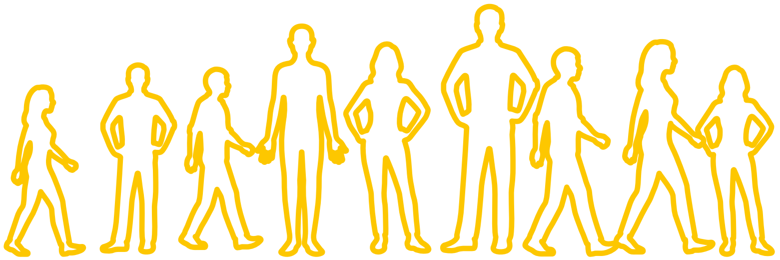
Two sample t-test

Independent Two-Sample t-Test

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
66.75	70.24	67.19	67.09	63.65	64.64	69.81	69.79	73.52	71.74

$$n_X = 10 \quad \bar{x} = 68.442 \quad s_X = 3.113$$

Height of 18 y/o in the US



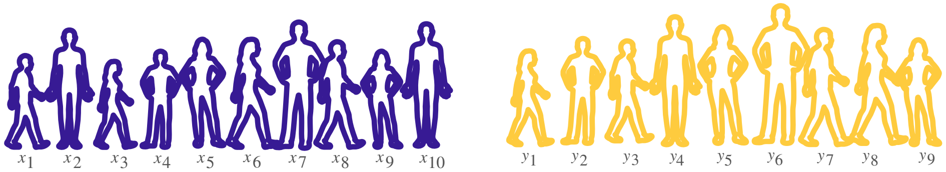
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
71.16	64.45	66.00	67.17	63.45	65.91	65.90	60.44	69.06

$$n_Y = 9 \quad \bar{y} = 65.949 \quad s_Y = 3.106$$

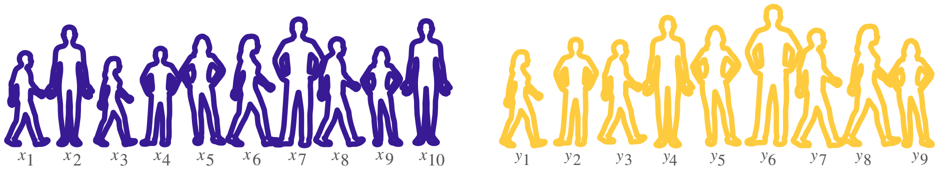
Height of 18 y/o in Argentina

$$\mu_{US} \neq \mu_{Ar}$$

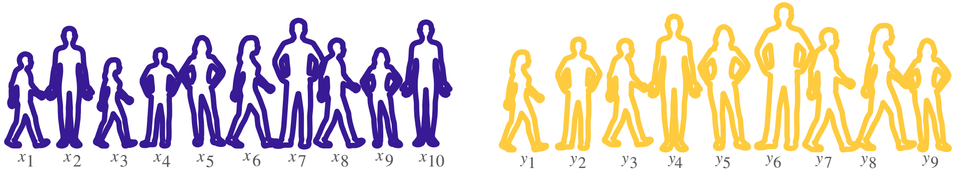
Independent Two-Sample t-Test: Hypothesis



$H_0 : \mu_{US} - \mu_{Ar} = 0$ vs. $H_1 : \mu_{US} - \mu_{Ar} > 0$



$H_0 : \mu_{US} - \mu_{Ar} = 0$ vs. $H_1 : \mu_{US} - \mu_{Ar} < 0$



$H_0 : \mu_{US} - \mu_{Ar} = 0$ vs. $H_1 : \mu_{US} - \mu_{Ar} \neq 0$

Independent Two-Sample t-Test: Assumptions

- All people in the sample from the two groups are different
- Each person in both samples are independent
- Populations are normally distributed

$$X \sim \mathcal{N}\left(\mu_{US}, \sigma_{US}^2\right) \qquad Y \sim \mathcal{N}\left(\mu_{Arg}, \sigma_{Arg}^2\right)$$

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$$

$$\overline{Y} = \frac{1}{9}\sum_{i=1}^{9} Y_i$$

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\quad - \quad , \frac{\quad}{\quad} + \frac{\quad}{\quad} \right)$$

Independent Two-Sample t-Test: Assumptions

- All people in the sample from the two groups are different
- Each person in both samples are independent
- Populations are normally distributed

$$X \sim \mathcal{N}\left(\mu_{US}, \sigma_{US}\right)$$

$$Y \sim \mathcal{N}\left(\mu_{Arg}, \sigma_{Arg}\right)$$

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$$

$$\overline{Y} = \frac{1}{9}\sum_{i=1}^{9} Y_i$$

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_{US} - \mu_{Arg}, \sqrt{\frac{\sigma_{US}^2}{10} + \frac{\sigma_{Ar}^2}{9}}\right)$$

Independent Two-Sample t-Test: Statistic

$X \sim \mathcal{N}\left(\mu_{US}, \sigma_{US}\right)$ $\quad$ $Y \sim \mathcal{N}\left(\mu_{Arg}, \sigma_{Arg}\right)$

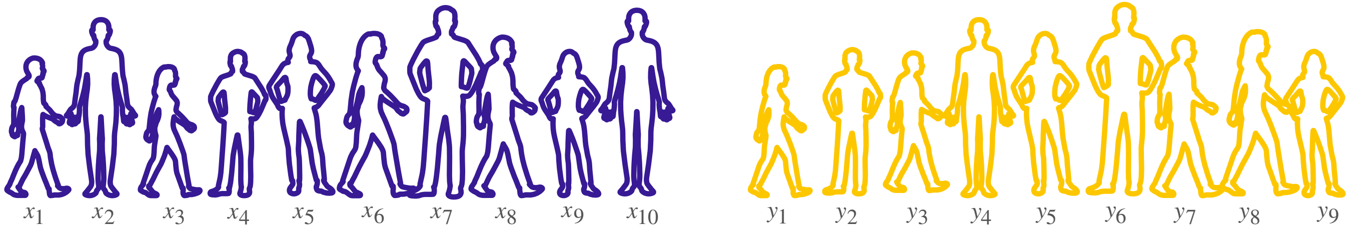
$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_{US} - \mu_{Arg}, \frac{\sigma_{US}^2}{10} + \frac{\sigma_{Ar}^2}{9}\right) \rightarrow \frac{(\bar{X} - \bar{Y}) - \left(\mu_{US} - \mu_{Arg}\right)}{\sqrt{\frac{\sigma_{US}^2}{10} + \frac{\sigma_{Ar}^2}{9}}} \sim \mathcal{N}(0,1^2)$$

You don't know σ_{US}, σ_{Ar} 😱 → Replace it with the sample standard deviation

$$T = \frac{(\bar{X} - \bar{Y}) - \left(\mu_{US} - \mu_{Arg}\right)}{\sqrt{\frac{S_X^2}{10} + \frac{S_Y^2}{9}}} \sim t_\nu$$

Degrees of freedom $= \dfrac{\left(\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_y}\right)^2}{\dfrac{\left(\frac{s_X^2}{n_X}\right)^2}{n_X - 1} + \dfrac{\left(\frac{s_Y^2}{n_Y}\right)^2}{n_Y - 1}}$

Independent Two-Sample t-Test



$n_X = 10$ $\quad \bar{x} = 68.442$ $\quad s_X = 3.113$

$n_Y = 9$ $\quad \bar{y} = 65.949$ $\quad s_Y = 3.106$

$$\text{Degrees of freedom} = \frac{\overbrace{\left(\frac{\ ^2}{\ } + \frac{\ ^2}{\ }\right)^2}^{16.8}}{\frac{\left(\frac{\ ^2}{\ }\right)^2}{\ -1} + \frac{\left(\frac{\ ^2}{\ }\right)^2}{\ -1}}$$

Independent Two-Sample t-Test: Statistic

$X \sim \mathcal{N}(\mu_{US}, \sigma_{US})$ $\quad$ $Y \sim \mathcal{N}(\mu_{Arg}, \sigma_{Arg})$

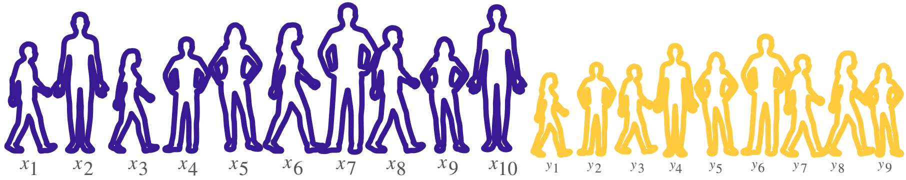
$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_{US} - \mu_{Arg}, \sqrt{\frac{\sigma_{US}^2}{10} + \frac{\sigma_{Ar}^2}{9}}\right) \rightarrow \frac{(\overline{X} - \overline{Y}) - (\mu_{US} - \mu_{Arg})}{\sqrt{\frac{\sigma_{US}^2}{10} + \frac{\sigma_{Ar}^2}{9}}} \sim \mathcal{N}(0,1)$$

You don't know σ_{US}, σ_{Ar} 😱 → Replace it with the sample standard deviation

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_{US} - \mu_{Arg})}{\sqrt{\frac{S_X^2}{10} + \frac{S_Y^2}{9}}} \sim t_{16.8}$$

Degrees of freedom $= \dfrac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_y}\right)^2}{\frac{\left(\frac{s_X^2}{n_X}\right)^2}{n_X - 1} + \frac{\left(\frac{s_Y^2}{n_Y}\right)^2}{n_Y - 1}}$

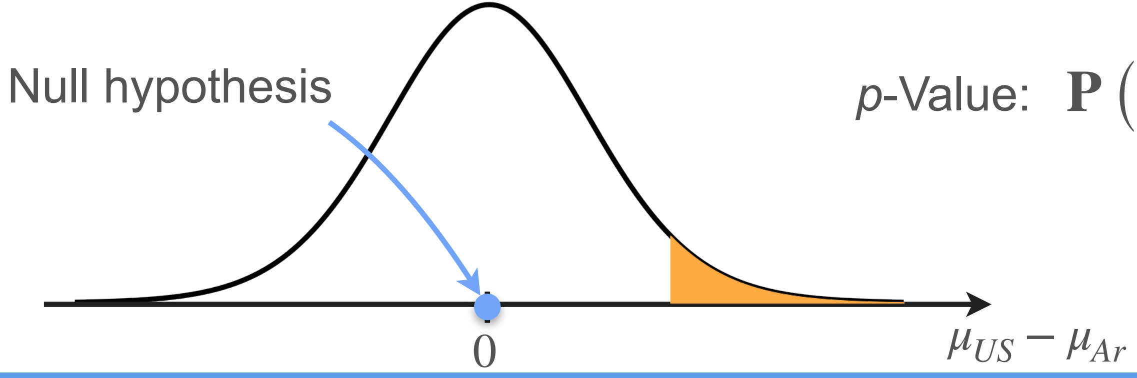
Independent Two-Sample t-Test: Right Tailed Test



$\bar{x} = 68.442$ $\quad n_X = 10$ $\quad \bar{y} = 65.949$ $\quad n_Y = 9$

$s_X = 3.113$ $\quad s_Y = 3.106$

$t = 1.7450$



$H_0 : \mu_{US} - \mu_{Ar} = 0$ vs. $H_1 : \mu_{US} - \mu_{Ar} > 0$

$\alpha = 0.05$

If H_0 is true: $T = \dfrac{(\bar{X} - \bar{Y}) - 0}{\sqrt{\dfrac{S_X^2}{10} + \dfrac{S_Y^2}{9}}} \sim t_{16.8}$

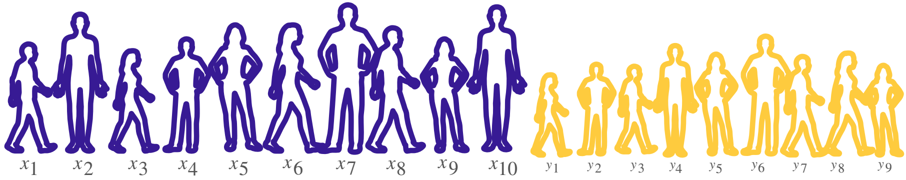
p-Value: $\mathbf{P}\left(T > \qquad | \mu_{US} - \mu_{Ar} = 0\right)$

$= 0.0495 < 0.05$

$\Rightarrow$ Reject H_0 (and accept H_1)
(with a 5% significance level)

DeepLearning.AI

Independent Two-Sample t-Test: Two Tailed Test



$\bar{x} = 68.442$ $\quad n_X = 10$ $\quad \bar{y} = 65.949$ $\quad n_Y = 9$

$s_X = 3.113$ $\quad s_Y = 3.106$

$t = 1.7450$

$H_0 : \mu_{US} - \mu_{Ar} = 0$ vs. $H_1 : \mu_{US} - \mu_{Ar} \neq 0$

$\alpha = 0.05$

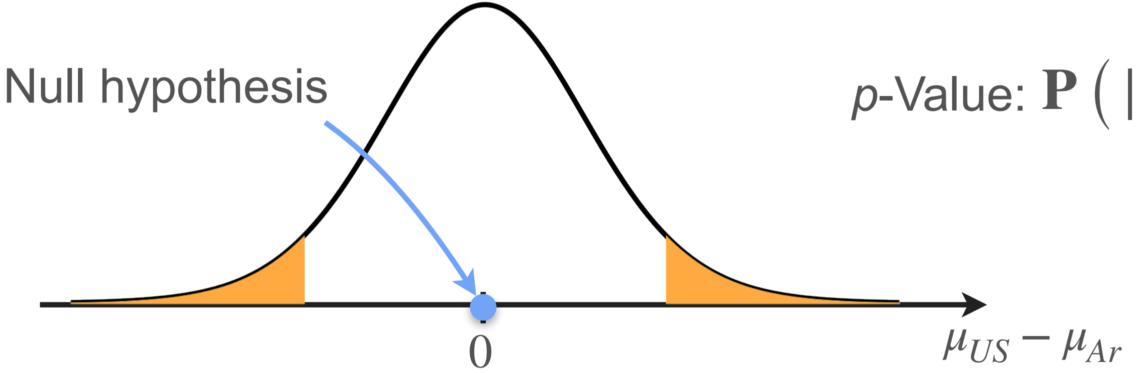
If H_0 is true: $T = \dfrac{(\overline{X} - \overline{Y}) - 0}{\sqrt{\dfrac{S_X^2}{10} + \dfrac{S_Y^2}{9}}} \sim t_{16.8}$

p-Value: $\mathbf{P}\left(|T| > \quad \middle| \mu_{US} - \mu_{Ar} = 0 \right)$

$= 0.0991 > 0.05$

$\Rightarrow$ Do not reject H_0

(with a 5% significance level)

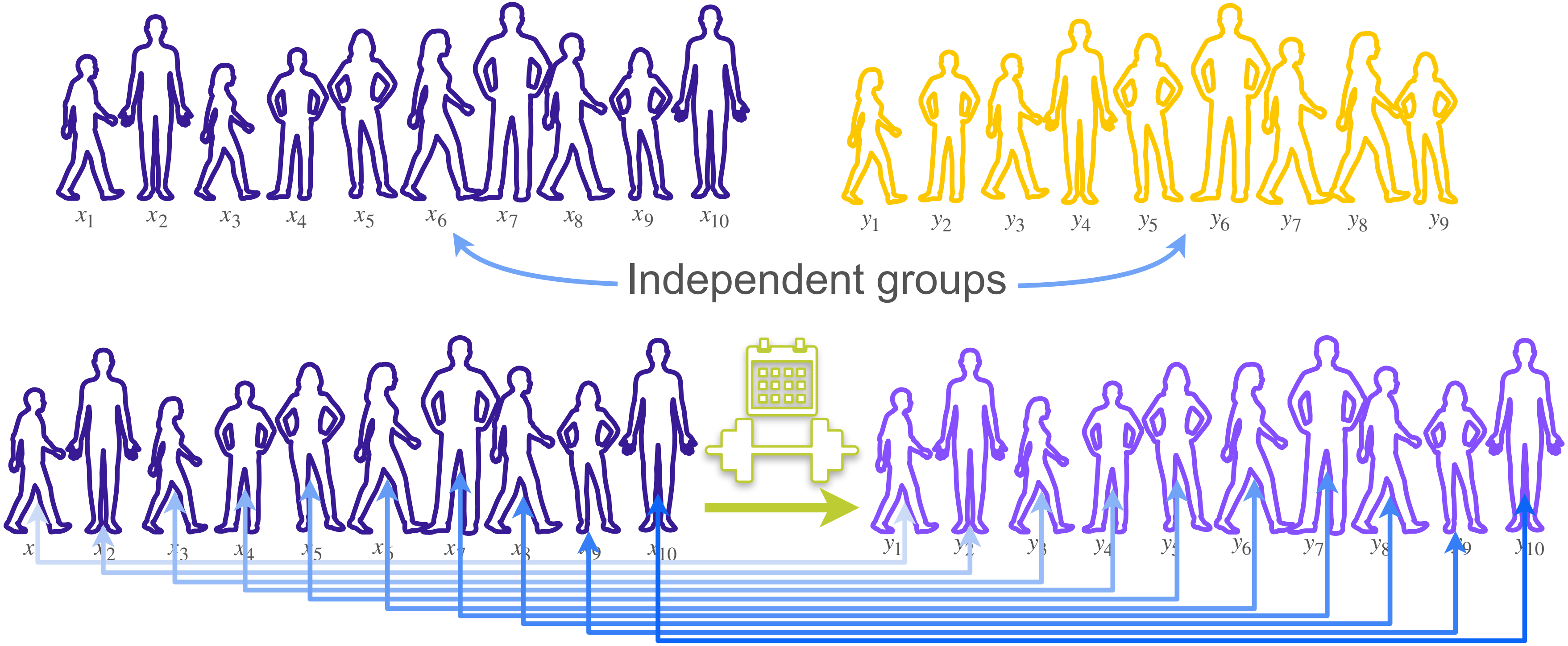


DeepLearning.AI

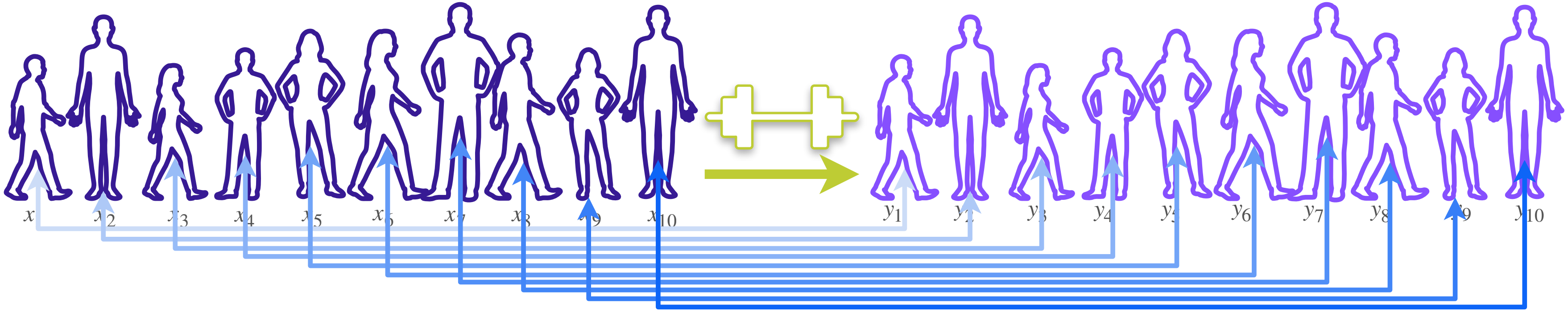
Hypothesis Testing

Paired t-test

Paired t-Test and Two-Sample t-Test



Paired *t*-Test: Statistic

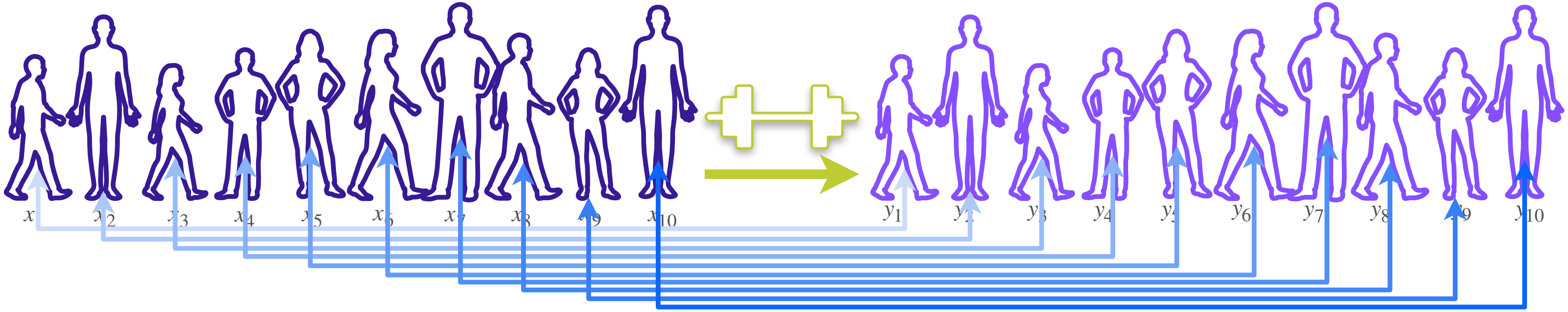


Now you're interested in the difference between pair of samples

$$\frac{(X_1 - Y_1)+(X_2 - Y_2)+(X_3 - Y_3)+(X_4 - Y_4)+(X_5 - Y_5)+(X_6 - Y_6)+(X_7 - Y_7)+(X_8 - Y_8)+(X_9 - Y_9)+(X_{10} - Y_{10})}{10}$$

$$\frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10}$$

Paired t-Test: Statistic

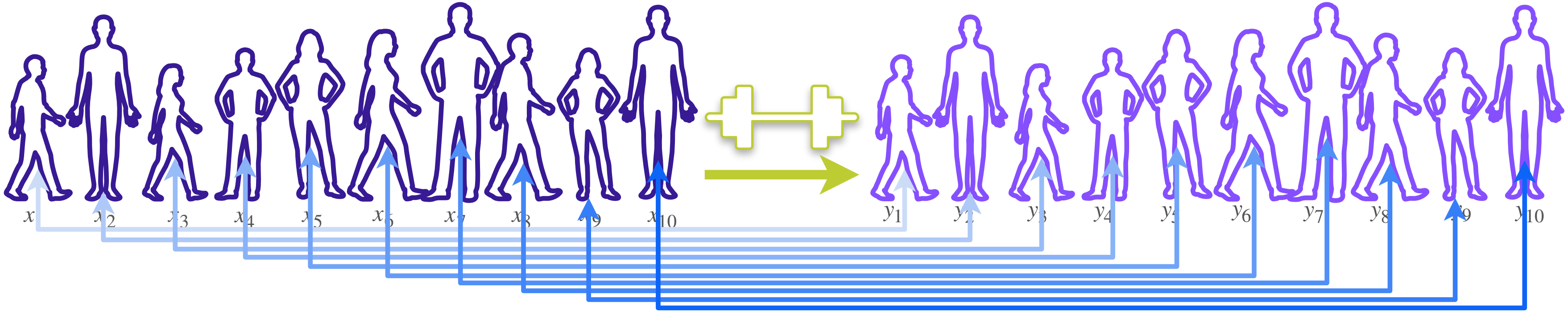


Now you're interested in the difference between pair of samples

$$\frac{(X_1 - Y_1)+(X_2 - Y_2)+(X_3 - Y_3)+(X_4 - Y_4)+(X_5 - Y_5)+(X_6 - Y_6)+(X_7 - Y_7)+(X_8 - Y_8)+(X_9 - Y_9)+(X_{10} - Y_{10})}{10}$$

$$\overline{D} = \frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10}$$

Paired t-Test: Statistic



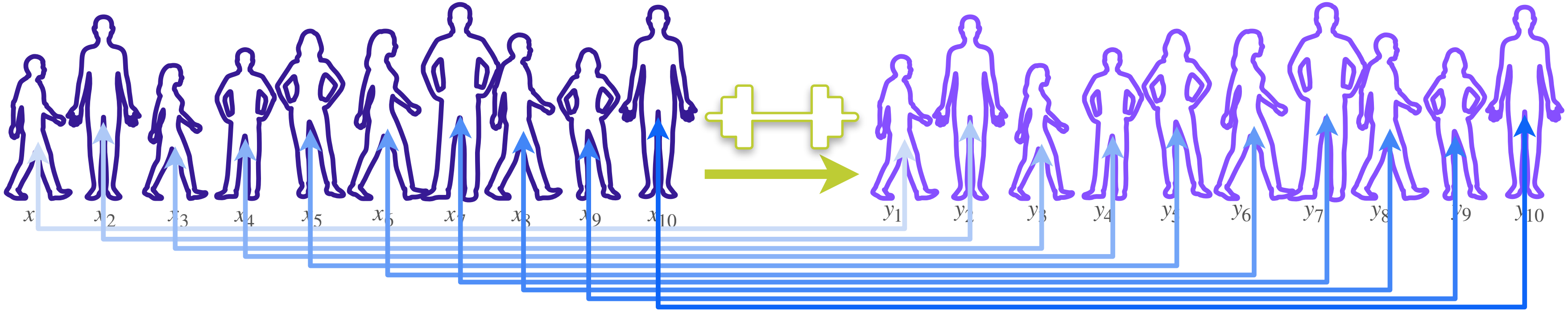
Now you’re interested in the difference between pair of samples

$$\overline{D} = \frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10}$$

If X_i, Y_i are gaussian $\Rightarrow D_i = X_i - Y_i$ is gaussian.

$$D_i \overset{i.i.d}{\sim} \mathcal{N}(\mu_D, \sigma_D^2)$$

Paired t-Test: Statistic



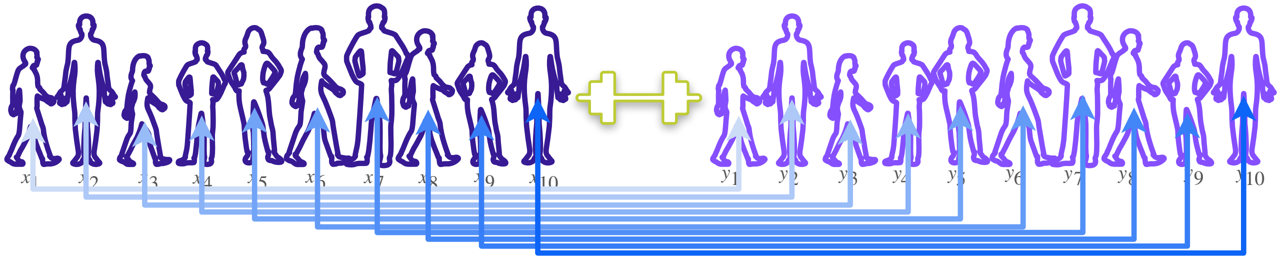
Now you're interested in the difference between pair of samples

$$\overline{D} = \frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10}$$

If X_i, Y_i are gaussian $\Rightarrow \boldsymbol{D_i = X_i - Y_i}$ is gaussian.

$$D_i \overset{i.i.d}{\sim} \mathcal{N}(\mu_D, \sigma_D^2)$$

Paired t-Test: Statistic



Now you're interested in the difference between pair of samples

$$\overline{D} = \frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10}$$

$$D_i = X_i - Y_i$$

$$D_i \overset{i.i.d}{\sim} \mathcal{N}(\mu_D, \sigma_D^2)$$

$$\frac{\overline{D} - \mu_D}{\sigma_D / \sqrt{10}} \sim \mathcal{N}(0, 1^2)$$

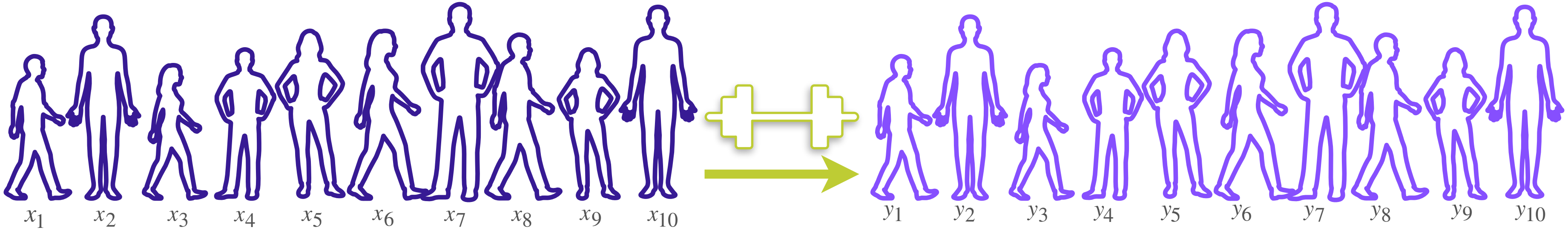
But σ_D is unknown

$$\Rightarrow \sigma_D \rightarrow S_D = \sqrt{\frac{\sum_{i=1}^{10} (D_i - \overline{D})^2}{10 - 1}} \Rightarrow \boxed{T = \frac{\overline{D} - 0}{S_D / \sqrt{10}} \sim t_{10-1}}$$

$H_0 : \mu_D = 0$

Test statistic

Paired t-Test: Observations



251.9	166.9	186.6	201.4	154.2	185.4	185.3	116.2	225.4	209.0

251.5	165.6	183.9	200.3	153.0	186.7	182.5	114.3	223.2	210.4

x_i 251.9

$-$

y_i 251.9

$=$

d_i	0.4	1.3	2.7	1.1	1.2	-1.3	2.8	1.9	2.2	-1.4

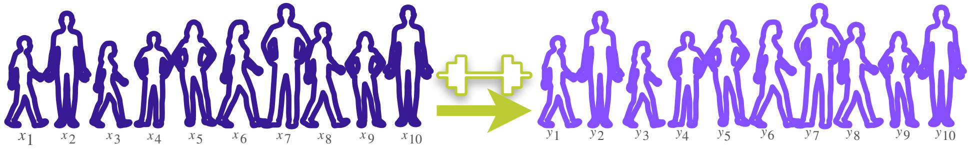
$$\bar{d} = 1.09$$

$$s_D = 1.485$$

$$t = \frac{-\ 0}{/\sqrt{10}}$$

$$= 2.321$$

Paired Two-Sample t-Test: Right Tailed Test

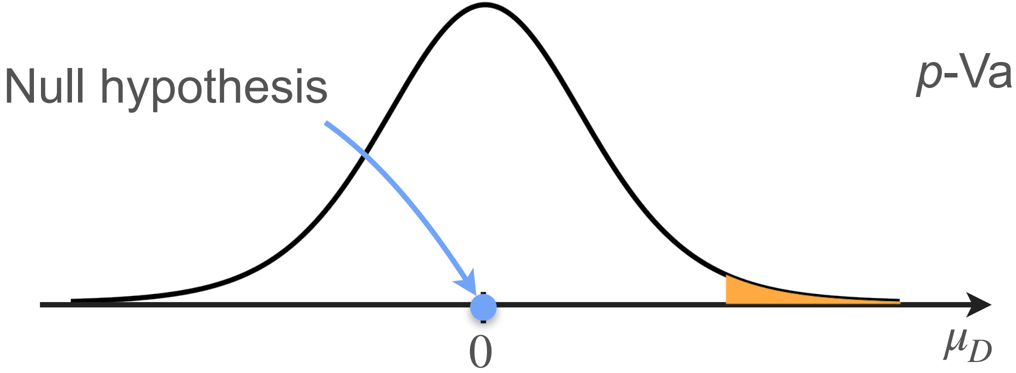


$$\underbrace{\bar{d} = 1.09 \qquad s_D = 1.485 \qquad n = 10}_{t \,=\, 2.321}$$

$H_0 : \mu_D = 0$ vs. $H_1 : \mu_D > 0$

$\alpha = 0.05$

If H_0 is true: $T = \dfrac{\overline{D} - 0}{S_D/\sqrt{10}} \sim t_{10-1}$



p-Value: $\mathbf{P}\left(T > \qquad \mid \mu_D = 0\right)$

$= 0.0227 < 0.05$

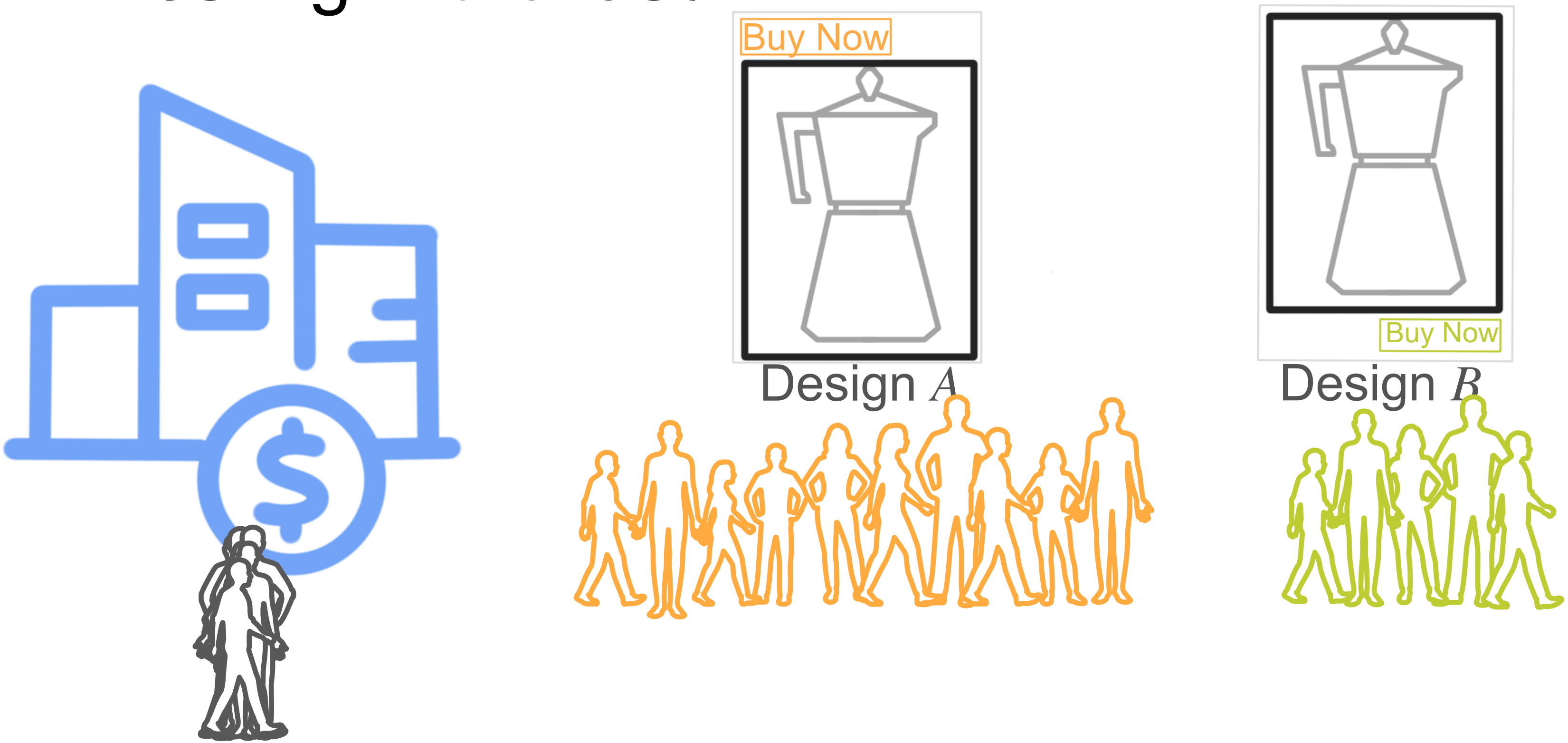
$\Rightarrow$ Reject H_0 (and accept H_1)
(with a 5% significance level)

DeepLearning.AI

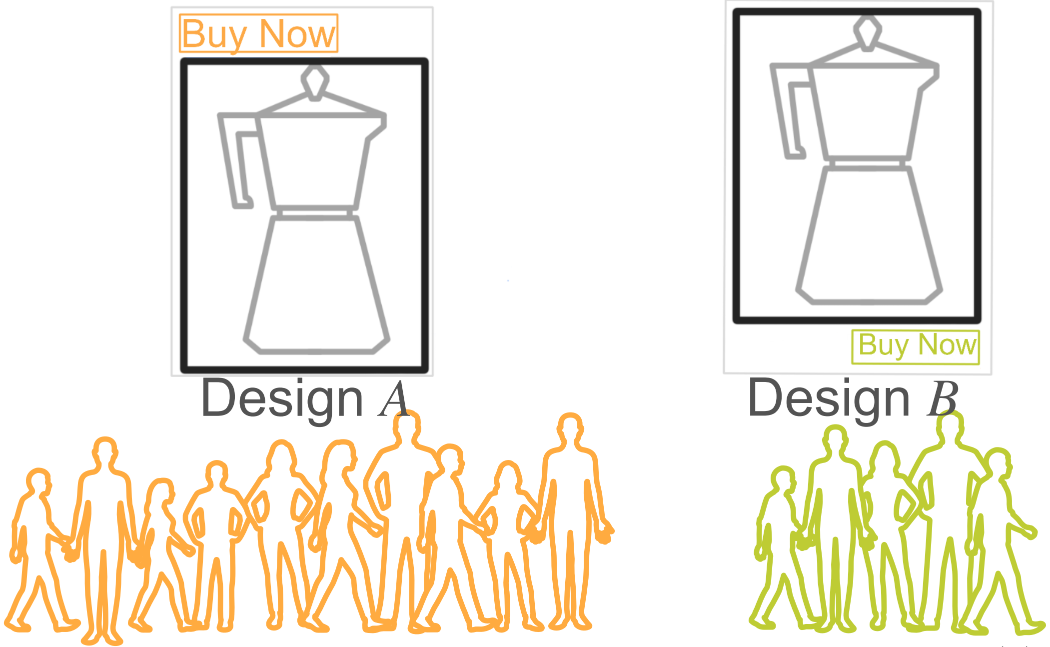
Hypothesis Testing

ML Application: A/B testing

A/B Testing: Purchase Amount



A/B Testing: Purchase Amount



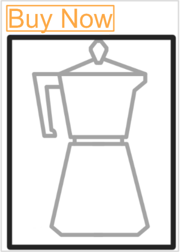
$n_A = 80$ $\quad \bar{x} = 50$

$s_A = 10$

$n_B = 20$ $\quad \bar{y} = 55$

$s_B = 15$

A/B Testing: Purchase Amount



Design A

Design B

$n_A = 80$ $\bar{x} = 50$ $n_B = 20$ $\bar{y} = 55$

$s_A = 10$ $s_B = 15$

$X \sim \mathcal{N}\left(\mu_A, \sigma_A^2\right)$ $Y \sim \mathcal{N}\left(\mu_B, \sigma_B^2\right)$

$H_0 : \mu_A - \mu_B = 0$ vs. $H_1 : \mu_A - \mu_B < 0$

$\alpha = 0.05$

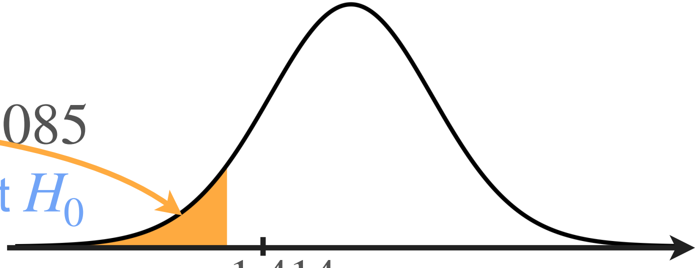
If H_0 is true: $T = \dfrac{\left(\overline{X} - \overline{Y}\right) - 0}{\sqrt{\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}}} \sim t_{23.38}$

$$t = \frac{(\quad - \quad) - 0}{\sqrt{\frac{\ ^2}{\ } + \frac{\ ^2}{\ }}}$$

-1.414

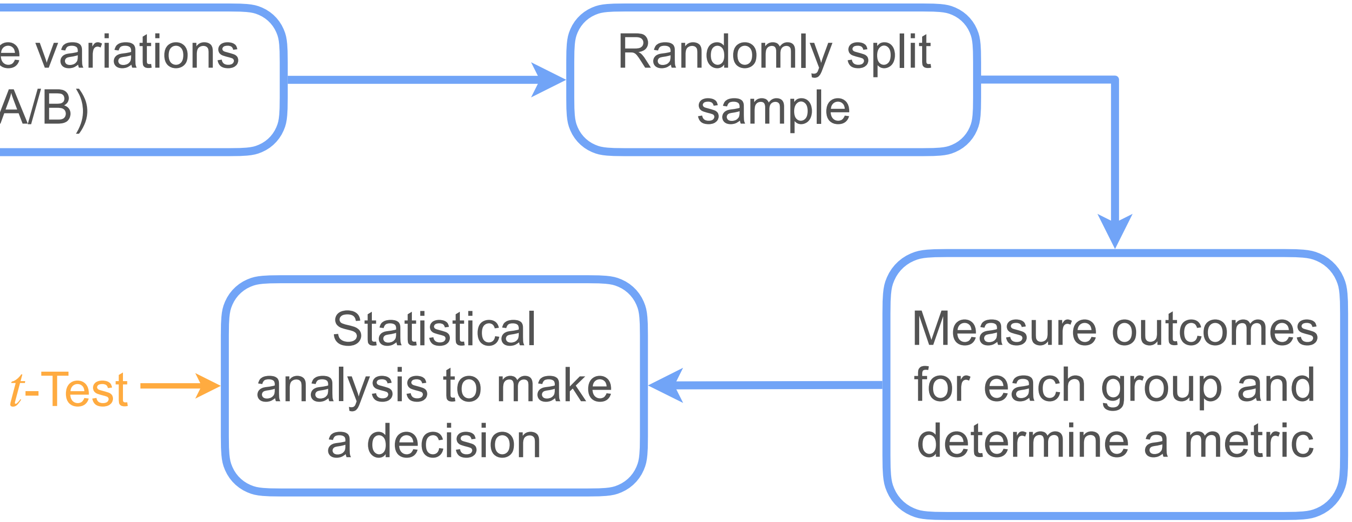
p-Value: 0.085

Don't reject H_0



A/B Testing and t-Tests

A/B testing is a methodology for comparing two variations (A/B)



A/B Testing: Conversion Rates

Design A

Design B

Higher conversion rates?

A/B Testing: Conversion Rates

$n_A = 80$ $x = 20$

$n_B = 20$ $y = 8$

A/B Testing: Conversion Rates

$n_A = 80$

$x = 20$

$n_B = 20$

$y = 8$

$X \sim Binomial(n_A, p_A)$ $\quad Y \sim Binomial(n_B, p_B)$

$H_0 : p_A - p_B = 0$ vs. $H_1 : p_A - p_B < 0$

$p_A =$ Conversion rate from Design A

$p_B =$ Conversion rate from Design B

$\alpha = 0.05$

A/B Testing: Conversion Rates

Statistic?

Law of large numbers

$$\frac{X}{n_A} \to p_A \quad\longrightarrow\quad \frac{X}{n_A} \sim \mathcal{N}\left(p_A, \frac{p_A(1-p_A)}{n_A}\right)$$

C.L.T.

$$\frac{Y}{n_B} \to p_B \quad\longrightarrow\quad \frac{Y}{n_B} \sim \mathcal{N}\left(p_B, \frac{p_B(1-p_B)}{n_B}\right)$$

A/B Testing: Conversion Rates

Statistic?

$$\frac{X}{n_A} \overset{a}{\sim} \mathcal{N}\left(p_A, \frac{p_A(1-p_A)}{n_A}\right)$$

$$\frac{Y}{n_B} \overset{a}{\sim} \mathcal{N}\left(p_B, \frac{p_B(1-p_B)}{n_B}\right)$$

$$\frac{X}{n_A} - \frac{Y}{n_B} \overset{a}{\sim} \mathcal{N}\left(\quad - \quad , \qquad\qquad + \qquad\qquad\right)$$

$$\frac{X}{n_A} - \frac{Y}{n_B} \to p_A - p_B$$

$$\Updownarrow$$

$$\frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - (p_A - p_B)}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}} \overset{a}{\sim} \mathcal{N}(0,1^2)$$

A/B Testing: Conversion Rates

If H_0 is true $\Rightarrow p_A = p_B = p$

$$\frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - (p \quad - p \quad)}{\sqrt{\boxed{\frac{p \ (1 - p \ )}{n_A} + \frac{p \ (1 - p \ )}{n_B}}}} \overset{a}{\sim} \mathcal{N}\left(0,1^2\right)$$

$$= p(1-p)\left(\frac{1}{n_A} + \frac{1}{n_B}\right) = p(1-p)(n_A + n_B)\frac{1}{n_An_B}$$

A/B Testing: Conversion Rates

If H_0 is true $\Rightarrow p_A = p_B = p$

$$\frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - (p - p)}{\sqrt{\frac{p(1-p)}{n_A} + \frac{p(1-p)}{n_B}}} \overset{a}{\sim} \mathcal{N}(0,1^2) \longrightarrow \frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - 0}{\sqrt{(n_A + n_B)p(1-p)}}\sqrt{n_A n_B} \overset{a}{\sim} \mathcal{N}(0,1^2)$$

But you don't know p

$\downarrow$

Replace it by estimation! $\hat{p} = \frac{X+Y}{n_A + n_B}$

Test statistic

$$\boxed{\frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - 0}{\sqrt{(X+Y)\left(1 - \frac{X+Y}{n_A + n_B}\right)}}\sqrt{n_A n_B} \overset{a}{\sim} \mathcal{N}(0,1^2)}$$

A/B Testing: Conversion Rates

Design A

$n_A = 80$

$x = 20$

$X \sim Binomial(n_A, p_A)$

Design B

$n_B = 20$

$y = 8$

$Y \sim Binomial(n_B, p_B)$

$H_0 : p_A - p_B = 0$ vs. $H_1 : p_A - p_B < 0$

$\alpha = 0.05$

If H_0 is true $\Rightarrow p_A = p_B = p$

$$Z = \frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - 0}{\sqrt{(X+Y)\left(1 - \frac{X+Y}{n_A+n_B}\right)}} \sqrt{n_A n_B} \sim \mathcal{N}(0, 1^2)$$

$$z = \frac{\left(\frac{}{} - \frac{}{}\right) - 0}{\sqrt{(+)\left(1 - \frac{+}{+}\right)}} \sqrt{}$$

$z = -1.336$

DeepLearning.AI

A/B Testing: Conversion Rates

Design A Design B

$n_A = 80$ $n_B = 20$

$x = 20$ $y = 8$

$X \sim Binomial(n_A, p_A)$ $Y \sim Binomial(n_B, p_B)$

$H_0 : p_A - p_B = 0$ vs. $H_1 : p_A - p_B < 0$

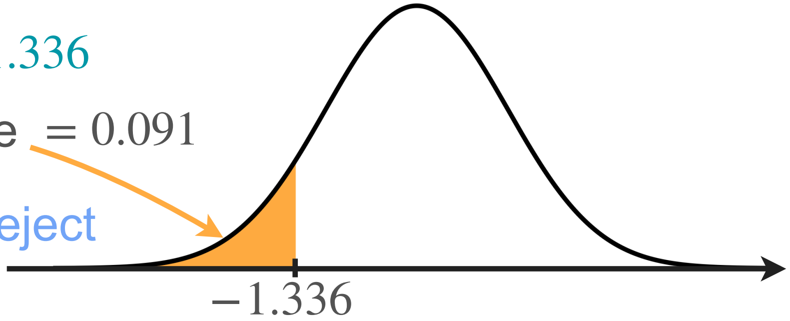
$\alpha = 0.05$ If H_0 is true $\Rightarrow p_A = p_B = p$

$$Z = \frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - 0}{\sqrt{(X+Y)\left(1 - \frac{X+Y}{n_A + n_B}\right)}} \sqrt{n_A n_B} \sim \mathcal{N}(0, 1^2)$$

$z = -1.336$

p-value $= 0.091$

Do not reject H_0



DeepLearning.AI

Hypothesis Testing

Conclusion