Day-68, Feb 5, 2025 (Magh 23, 2081)

1. Type I and Type II errors
2. Right-tailed, Left-Tailed and Two-Tailed Tests
3. p-value
4. Critical value

$\alpha \rightarrow$ Accept the mistakes.

p-value > α → Reject H_0 ⟸ Critical value > calc. Absolute value

p-value < α → Accept H_0 ⟸ Critical value < calc. Absolute value

$\mu < \mu_0 \rightarrow$ Left-tailed

$\mu > \mu_0 \rightarrow$ Right-tailed

$\mu \neq \mu_0 \rightarrow$ Two-tailed

This detailed content covers the essential steps and concepts for hypothesis testing, particularly when dealing with population means. Here's a streamlined explanation with the key points highlighted:

Hypothesis Testing for Population Mean

Types of Tests

Right-Tailed Test: H_0: $\mu = 66.7$, H_1: $\mu > 66.7$

Use when testing if the mean has increased.

Left-Tailed Test: H_0: $\mu = 66.7$,H_1: $\mu < 66.7$

Use when testing if the mean has decreased.

Two-Tailed Test: H_0: $\mu = 66.7$, H_1: $\mu \neq 66.7$

Use when testing for any change in the mean.

Test Statistic

The sample mean (68.442 in this example) is used as the test statistic.

The observed statistic is the calculated mean from the sample data.

Errors

Type I Error: Incorrectly rejecting H_0 when it is true.

Type II Error: Failing to reject H_0 when it is false.

P-Value Concept

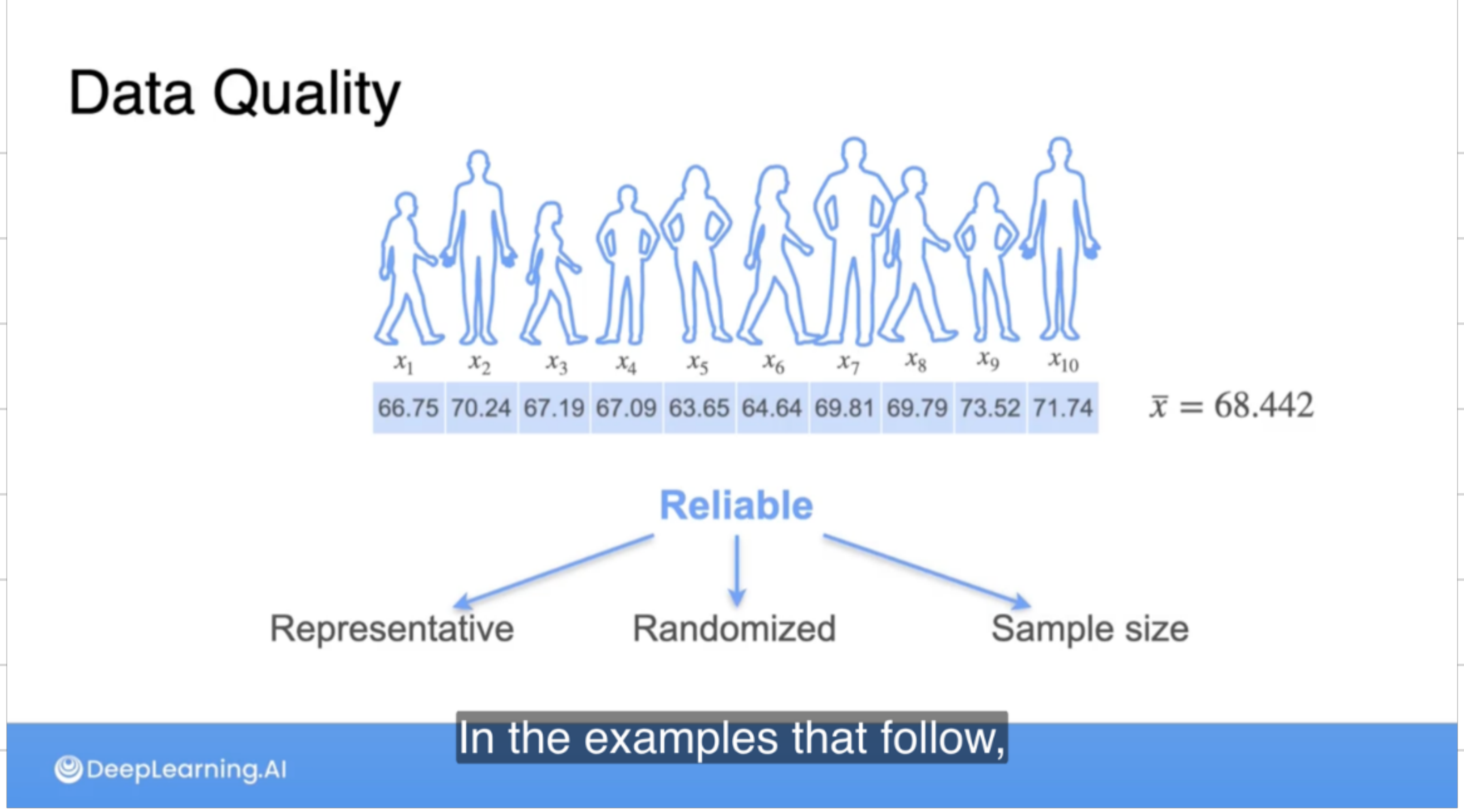
Indicates the probability of obtaining a sample mean as extreme as the observed one if H_0 is true.

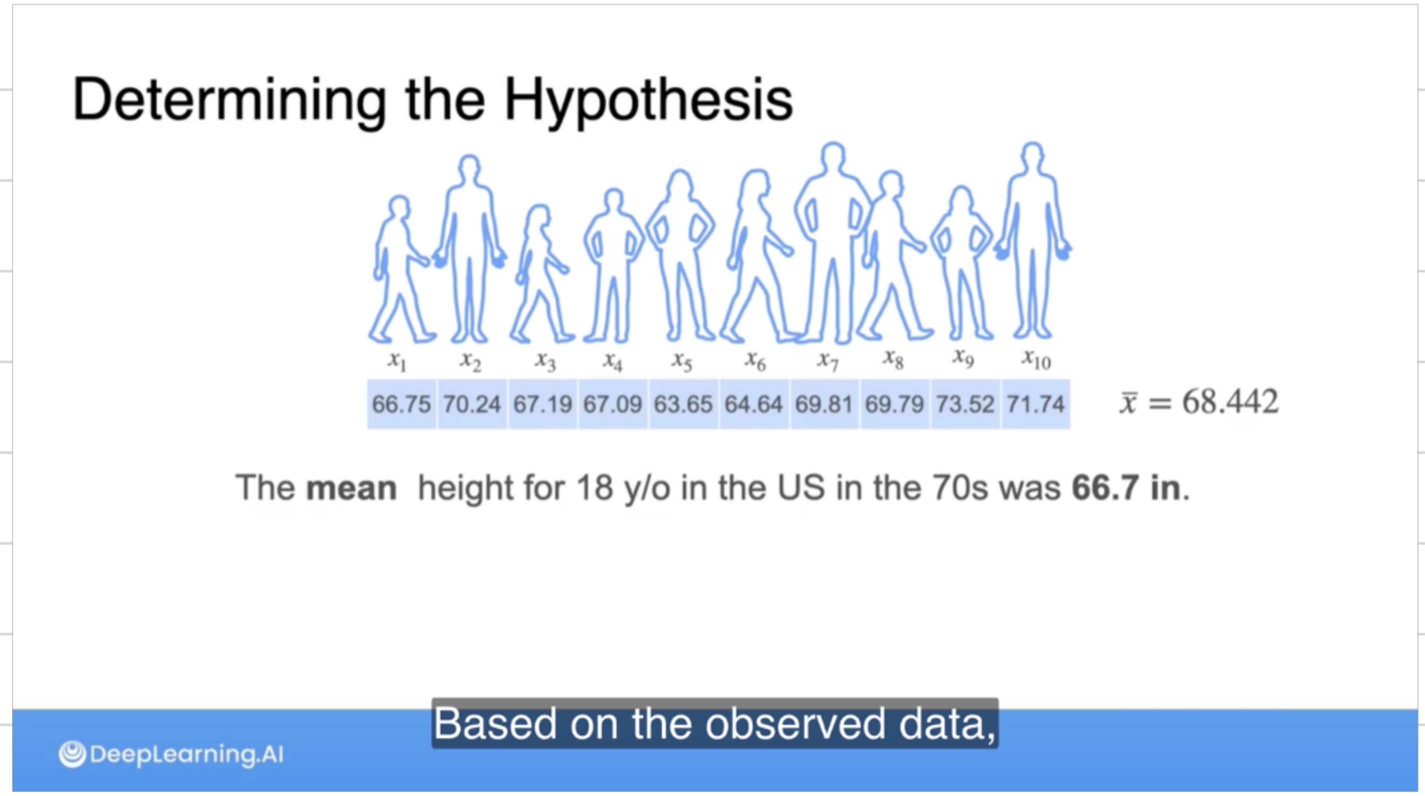
A small p-value (< significance level, typically 0.05) leads to rejecting H_0.

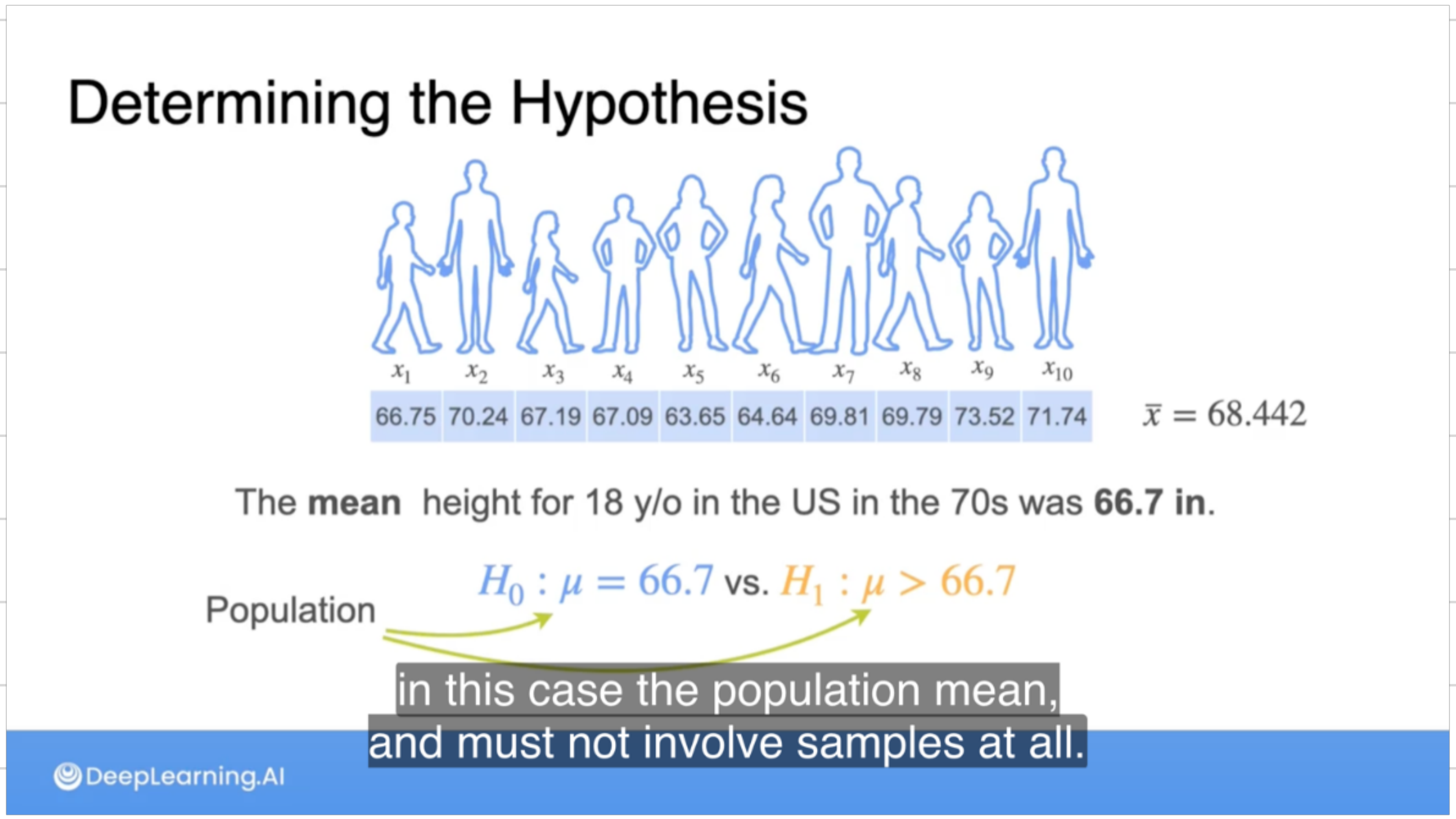
Important Data Considerations

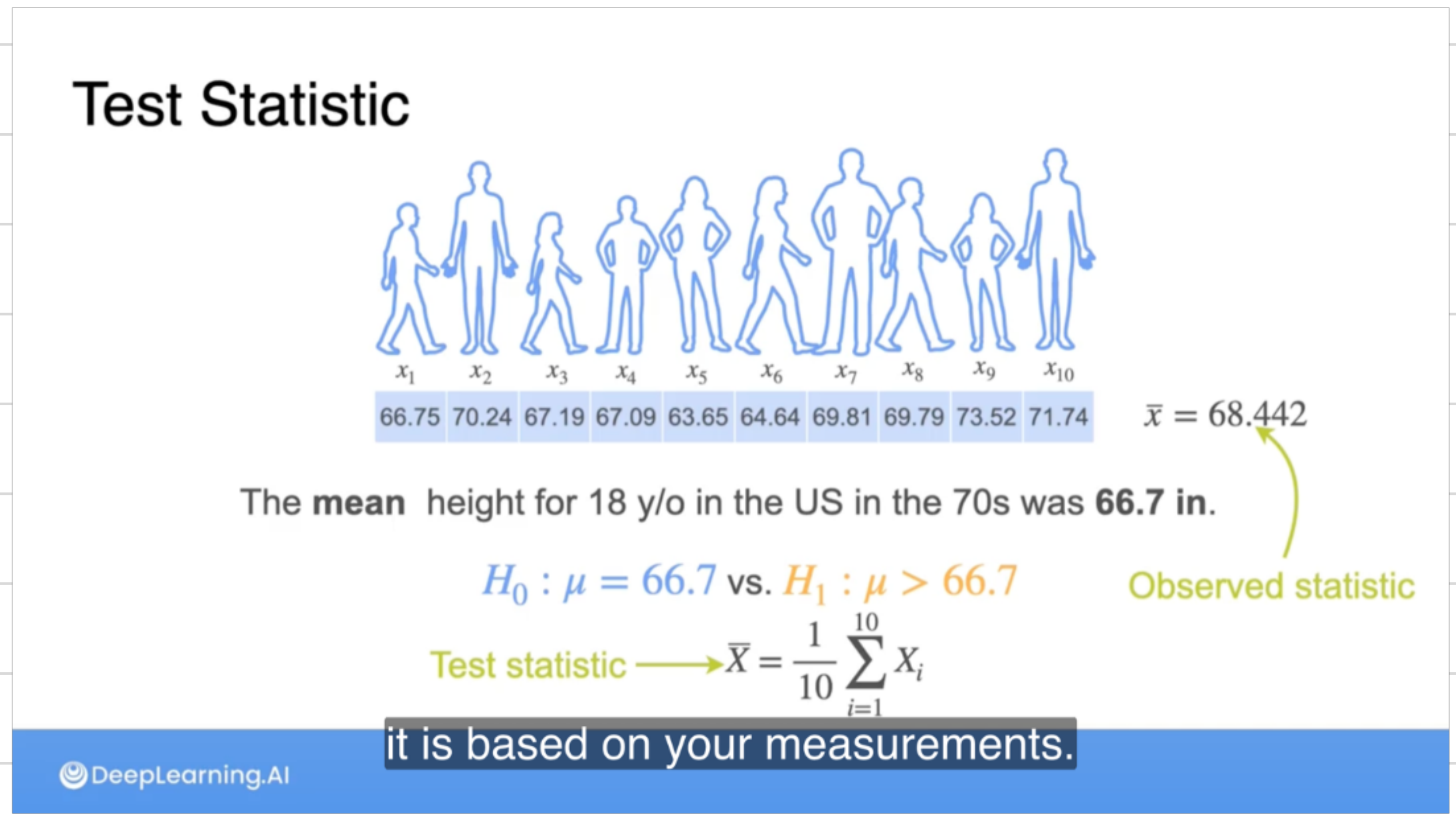
Sample Representativeness: Data must represent the entire population without bias.

Sample Size: A minimum of 30 samples is recommended for reliable conclusions.


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
In the examples that follow,
DeepLearning.AI


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**.
Based on the observed data,
DeepLearning.AI


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**.
Population
$H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$
in this case the population mean,
and must not involve samples at all.
DeepLearning.AI


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 $\longrightarrow \bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$
it is based on your measurements.
DeepLearning.AI

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

it is based on your measurements.

DeepLearning.AI

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: $T(X)$ $\quad X = (X_1, \ldots, X_n)$

Information about the population parameter under study

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

Not unique!

DeepLearning.AI

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

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

DeepLearning.AI

Example: Heights

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

3 questions $\qquad$ **3 sets of hypothesis**

Right-Tailed Test $\longrightarrow H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

This one is called a right-tailed test,
because the alternative

DeepLearning.AI

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**

Right-Tailed Test → $H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

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

This kind of hypothesis is called a left-tailed test,

DeepLearning.AI

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**

Right-Tailed Test → $H_0 : \mu = 66.7$ vs. $H_1 : \mu > 66.7$

Left Tailed Test → $H_0 : \mu = 66.7$ vs. $H_1 : \mu < 66.7$

Two-Tailed Test → $H_0 : \mu = 66.7$ vs. $H_1 : \mu \neq 66.7$

And the null hypothesis is the same.

DeepLearning.AI

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

Null hypothesis

Alternative hypothesis

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

DeepLearning.AI

What are the two errors we can have here?

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

Null hypothesis

Alternative hypothesis

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

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

DeepLearning.AI

the true value was actually 66.7.

Example: Heights

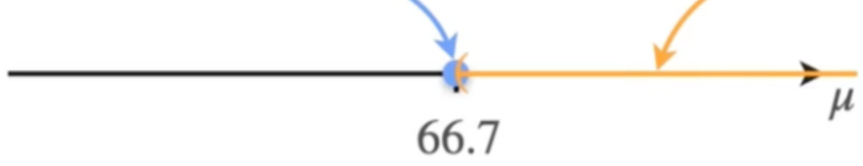
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\overline{X}$ Test statistic

Right-tailed test → $H_0 : \mu = 66.7$ 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 $\mu > 66.7$

Question

If the null hypothesis H_0 states that the mean height for 18-year-olds in the US in the 1970s was 66.7 inches or $H_0 : \mu = 66.7$, and the alternative hypothesis H_1 is that the mean height is less than 66.7 inches or $H_0 : \mu < 66.7$, which statement correctly identifies a type 1 error ad a type II error in this context?

- (selected) Type I error: Determining that $\mu < 66.7$ when the population mean did not change.

  Type II error: Determining that $\mu = 66.7in$ when $\mu < 66.7$ is true.
- Type I error: Determining that $\mu < 66.7in$ when $\mu \geq 66.7$ is true.

  Type II error: Determining that $\mu = 66.7in$ when $\mu \geq 66.7$ is true.
- Type I error: Determining that $\mu = 66.7in$ when $\mu < 66.7$ is true.

  Type II error: Determining that $\mu < 66.7$ when the population mean did not change.

Correct
Great job! A type I error occurs when we incorrectly reject the null hypothesis when it is actually true, while a type II error occurs when we fail to reject the null hypothesis when it is actually false.

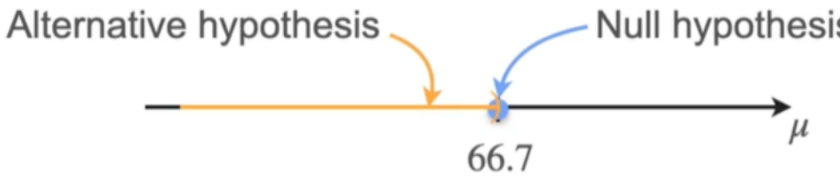
Skip | Continue

Example: Heights

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

$\overline{X}$

Left tailed test → $H_0 : \mu = 66.7$ vs. $H_1 : \mu < 66.7$



Type I error:

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

Type II error:

In this case a type I error happens when there was no change in the population

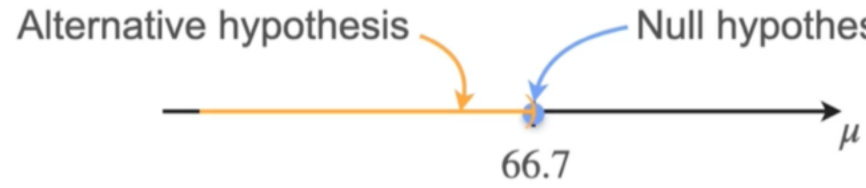
DeepLearning.AI

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



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 $\mu < 66.7$

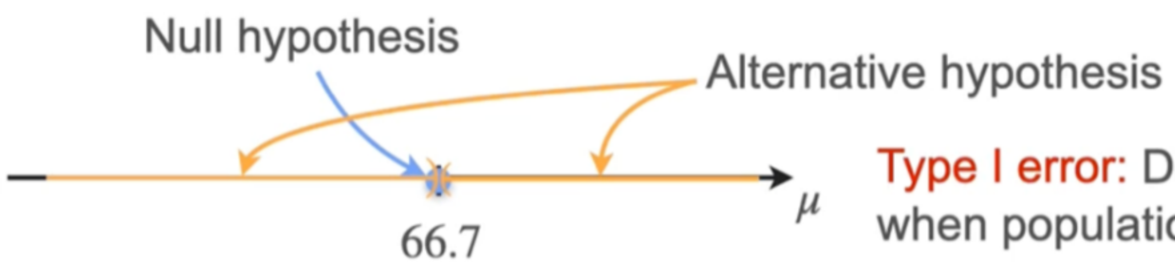
DeepLearning.AI

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. $H_1 : \mu \neq 66.7$

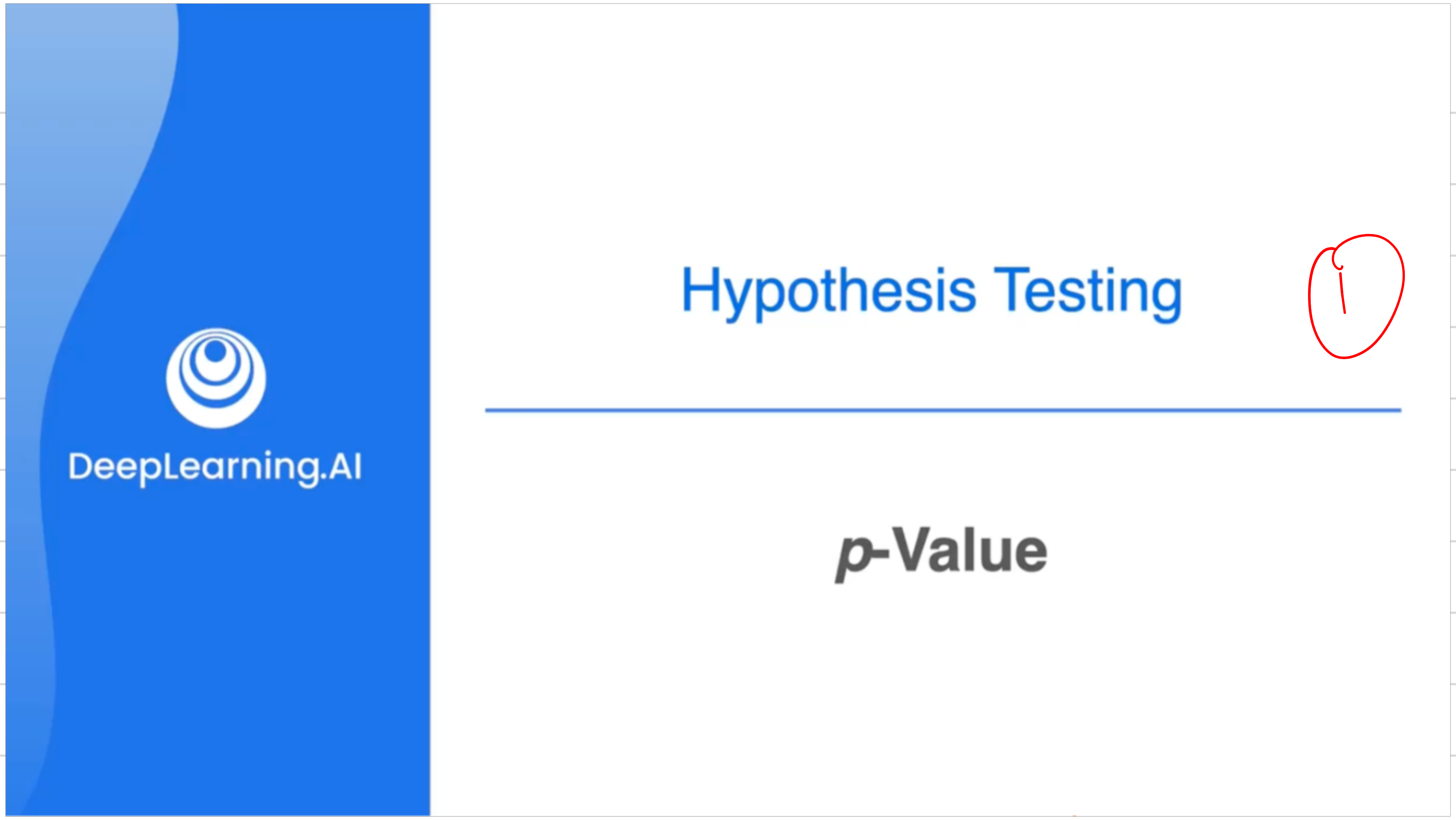


Type I error: Determine $\mu \neq 66.7$, when population mean did not change

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

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

DeepLearning.AI



1

2

p-value works against the Ho (Null Hypothesis)

A 5% alpha ($\alpha = 0.05$) means you are willing to accept a 5% chance of making a Type I error, also known as a false positive. In the context of hypothesis testing:

Type I error (False Positive): This is when you reject the null hypothesis when it's actually true. Imagine a medical test for a disease. A false positive means the test says you have the disease when you actually don't. With a 5% alpha, you're accepting a 5% risk that your study might conclude there's an effect (e.g., a drug works) when in reality there isn't.

p-value is

The p-value is the probability of obtaining results as extreme as, or more extreme than, the ones observed, assuming the null hypothesis is true. It's a measure of how surprising your data are if the null hypothesis is actually correct. It works agains the null hypothesis.

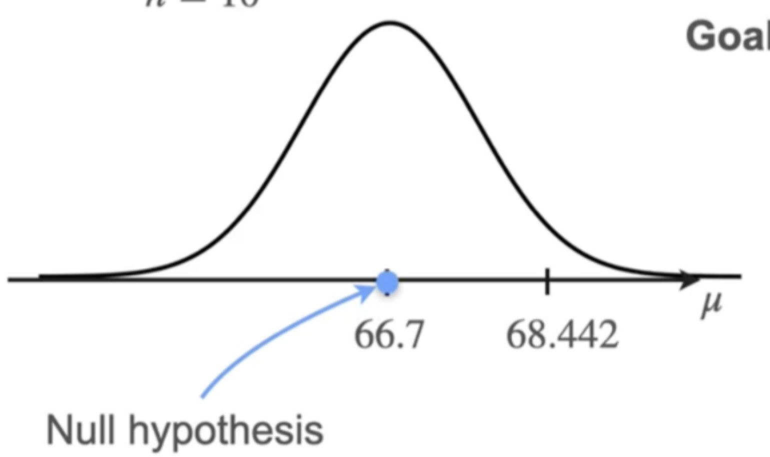
Right-Tailed Test for Gaussian Data (Known σ)

3

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

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

DeepLearning.AI

Right-Tailed Test for Gaussian Data (Known σ)

4

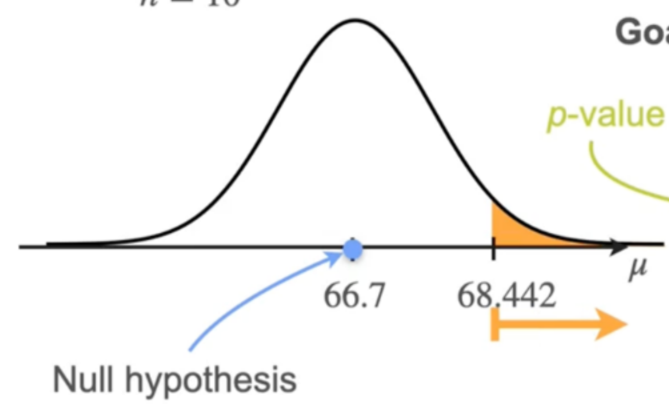
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$
$n = 10$

$\bar{x} = 68.442$

$H_0 : \mu = 66.7$ 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

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

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

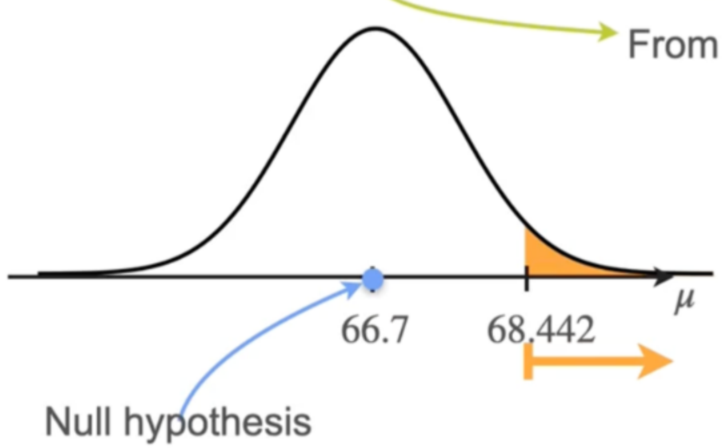
DeepLearning.AI

P-Values

5

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

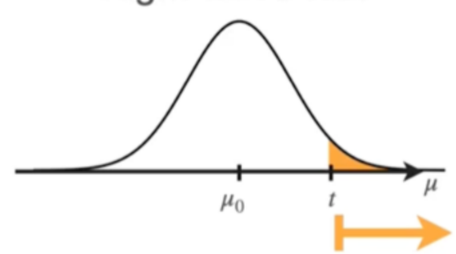
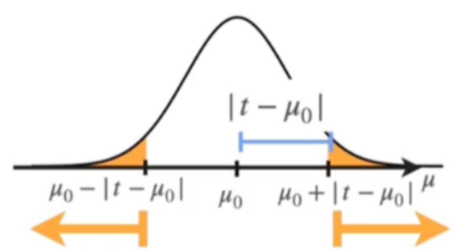
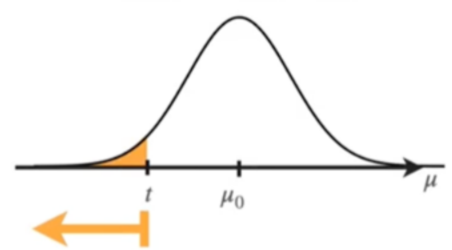
DeepLearning.AI

p-values

6

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 t: observed statistic H_0: $\mu = \mu_0$

Right-tailed test	Two-tailed test	Left-tailed test
		
$\mathbf{P}(T(X) > t \mid H_0)$	$\mathbf{P}(\lvert T(X) - \mu_0 \rvert > \lvert t - \mu_0 \rvert \mid H_0)$	$\mathbf{P}(T(X) < t \mid H_0)$

DeepLearning.AI

11

Right-Tailed Test Using the Z Statistic

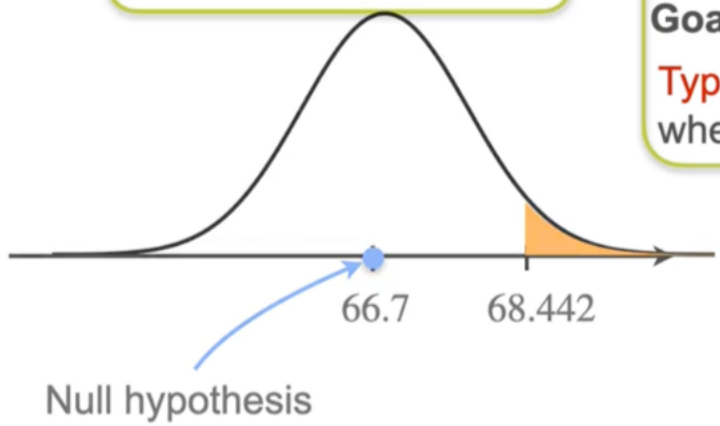
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

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

$H_0 : \mu = 66.7$ 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



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

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

DeepLearning.AI

12

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

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

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

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

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

DeepLearning.AI

13

Right-Tailed Test Using the Z Statistic

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

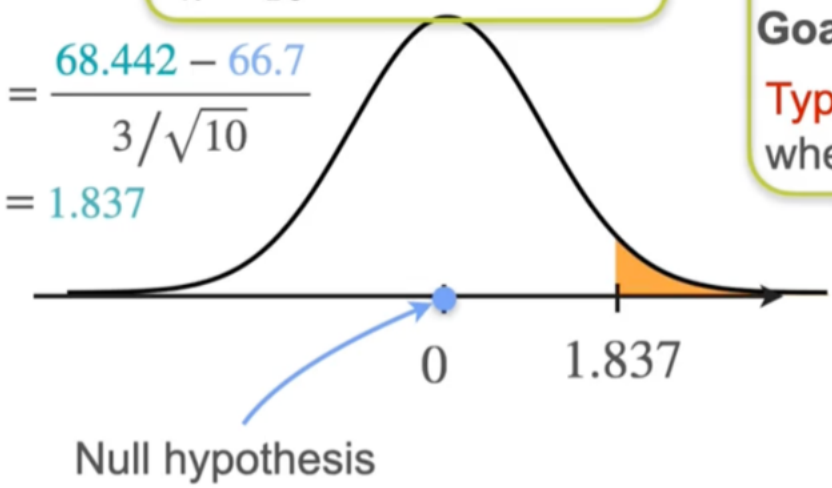
$\sigma = 3 \quad \bar{x} = 68.442$
$n = 10$

$H_0 : \mu = 66.7$ 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

$$z = \frac{68.442 - 66.7}{3/\sqrt{10}} = 1.837$$



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

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

DeepLearning.AI

Sources: Probability and Statistics for Data Science and Machine Learning. Coursera

Two-Tailed Test for Gaussian Data (Known σ)

7

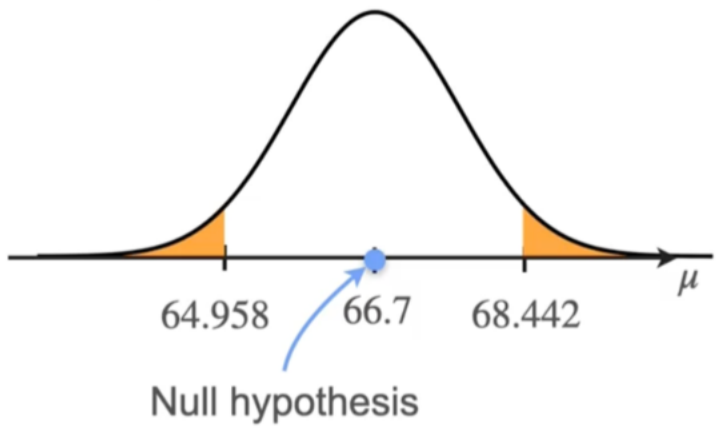
The **mean** height for 18 y/o in the US in the 70s was **66.7 in**.

$\sigma = 3$

$n = 10$

$\bar{x} = 68.442$

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



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

$$\mathbf{P}\left(\left|\bar{X} - 66.7\right| > \left|68.442 - 66.7\right| \,\middle|\, \mu = 66.7 \right) = 0.0663$$

DeepLearning.AI

Left-Tailed Test for Gaussian Data (Known σ)

8

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

$\sigma = 3$

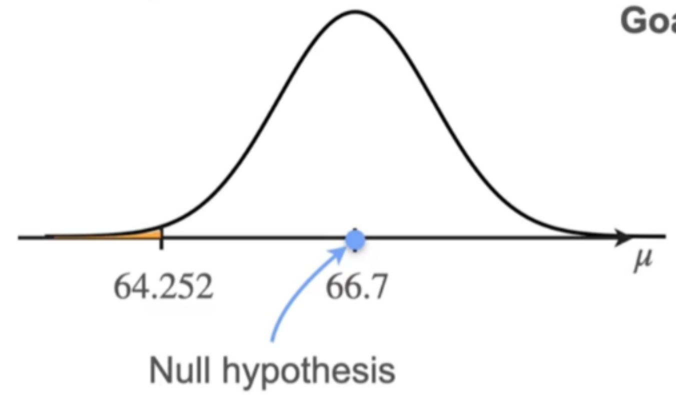
$n = 10$

$\bar{x} = 64.252$

$H_0 : \mu = 66.7$ 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)

DeepLearning.AI

Tests Using the Z-Statistic

9

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

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

Applying standardization, you can write equivalent tests using the

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

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

DeepLearning.AI

Right-Tailed Test Using the Z Statistic

10

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

$\sigma = 3$

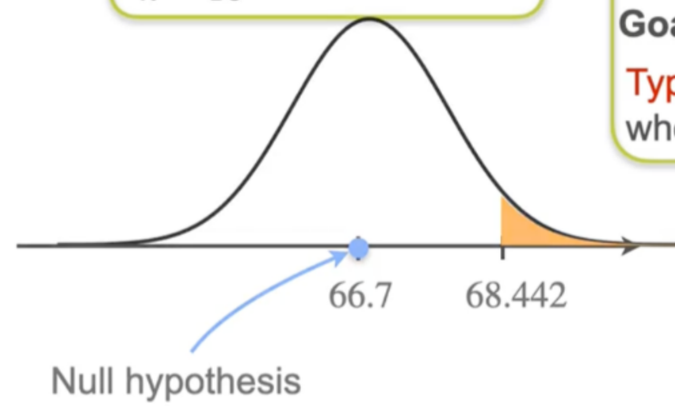
$n = 10$

$\bar{x} = 68.442$

$H_0 : \mu = 66.7$ 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)

DeepLearning.AI

Null Hypothesis (H_0): The population mean is 66.7 inches. We test if the sample mean is significantly different from this value.

Test Setup: We assume the standard deviation (σ) is known (3 inches) and a sample size of 10. The sample mean is 68.442 inches. If H_0 is true, the sample mean follows a Gaussian distribution with a mean of 66.7 and a variance 3^2/10.

Type 1 Error: A type 1 error occurs if we reject H_0 when it is true. The significance level (α) is set to 0.05, meaning we tolerate a 5% probability of committing a type 1 error.

P-value: The p-value is the probability of observing a sample mean as extreme as the observed mean (68.442) under the assumption that H_0 is true. If the p-value is less than α, we reject H_O.

Right-Tail Test: For the right-tail test, we look for the probability of observing a sample mean greater than 68.442. If the p-value is 0.332 (greater than 0.05), we do not reject H_0.

Two-Tailed Test: In this case, we check the probability of observing a sample mean farther from 66.7 than 1.742 units in both directions. The p-value for this test is 0.0663, which is greater than 0.05, so we fail to reject H_0.

Left-Tail Test: If the sample mean was lower (e.g., 64.252), the p-value would be smaller (0.0049), allowing us to reject H_0 and conclude that the population mean has decreased.

Z-statistic: An alternative approach is to standardize the sample mean using the z-statistic. The p-value remains the same, but the test uses the standard normal distribution to assess how likely the sample mean is under
H
0
H
0

.

This text discusses hypothesis testing, p-values, and the process of deciding whether to reject a null hypothesis (H_0). It covers different types of tests, including right-tail, left-tail, and two-tailed tests, and explains how the p-value is used to assess the likelihood of observing an extreme sample mean under the assumption that H_0 is true.

1

Key Concepts:

1. **Null Hypothesis (H_0)**: The assumption that the population mean is 66.7 inches. The test examines whether the observed sample mean deviates significantly from this value.
2. **Type 1 Error**: Incorrectly rejecting H_0 when it is true (e.g., claiming the population mean has changed when it has not).
3. **Significance Level (α)**: The probability threshold for making a Type 1 error, typically set at 0.05. If the p-value is smaller than α, H_0 is rejected.
4. **p-value**: The probability of obtaining a test statistic as extreme as the observed value (or more extreme), assuming H_0 is true. A small p-value indicates that the observed sample is unlikely under H_0, suggesting H_0 may be rejected.

Types of Tests:

- **Right-Tail Test**: Used when we hypothesize that the population mean is greater than a specified value. The p-value is the probability of obtaining a sample mean greater than or equal to the observed value under H_0.
- **Left-Tail Test**: Used when we hypothesize that the population mean is less than a specified value. The p-value is the probability of obtaining a sample mean smaller than or equal to the observed value under H_0.
- **Two-Tailed Test**: Used when we hypothesize that the population mean is different from a specified value, without specifying a direction. The p-value is the probability of obtaining a sample mean that deviates from the null hypothesis value in either direction.

Decision Rule:

2

- If the p-value < α, reject H_0.
- If the p-value ≥ α, do not reject H_0.

In the example, the p-value for a **right-tail test** with an observed sample mean of 68.442 is found to be 0.332, which is greater than the significance level (0.05), so H_0 is not rejected. For a **two-tailed test**, the p-value is 0.0663, which is also greater than 0.05, so again H_0 is not rejected. Finally, for a **left-tail test**, with an observed mean of 64.252, the p-value is 0.0049, which is smaller than 0.05, so H_0 is rejected, suggesting a decrease in the population mean.

p-Value Calculation:

For each type of test (right-tail, left-tail, and two-tailed), the p-value is calculated using the test statistic (e.g., z-statistic) and the corresponding probability from the standard normal distribution.

The **z-statistic** is the standardized version of the sample mean, allowing you to use the standard normal distribution (mean = 0, variance = 1) to calculate the p-value. The z-statistic formula is:

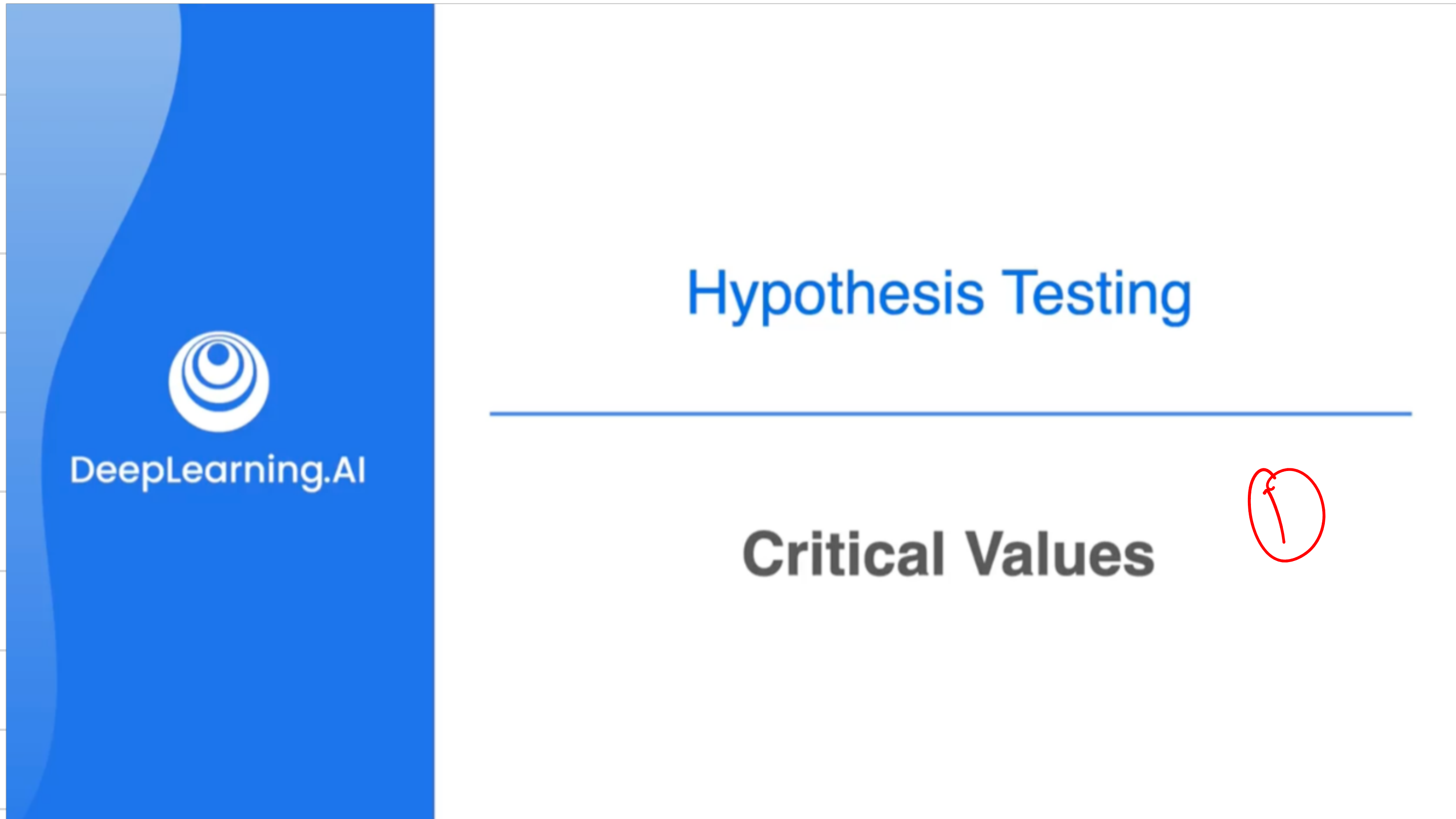
$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Where:

- $\bar{x}$ = observed sample mean,
- μ_0 = population mean under H_0,
- σ = population standard deviation,
- n = sample size.

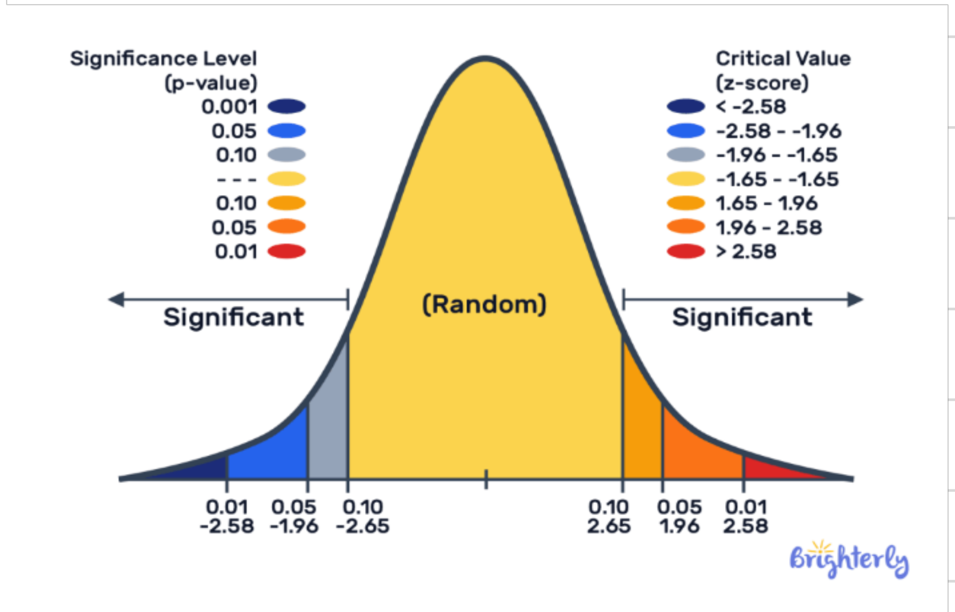
In this case, for the right-tail test, the z-statistic for the observed sample mean (68.442) is calculated as 1.837.

Conclusion:
Right-Tail Test: p-value = 0.332 → Do not reject H_0.
Two-Tailed Test: p-value = 0.0663 → Do not reject H_0.
Left-Tail Test: p-value = 0.0049 → Reject H_0.
This is a typical application of hypothesis testing, where we use p-values to evaluate how likely or unlikely our sample mean is, given the assumption that the null hypothesis is true.



1

Critical values are like the boundaries or cut-off points on a statistical distribution that help us decide whether to reject the null hypothesis. They are determined by the chosen significance level (alpha) and the type of test being conducted.



2

In hypothesis testing, the p-value and critical value are two ways to decide whether to reject the null hypothesis (H_0):

P-value: Represents the probability of getting an observed result (or more extreme) under H_0. If the p-value is less than the significance level (α), reject H_0.

Critical value: A predefined threshold that separates the rejection region from the non-rejection region. If the observed statistic is more extreme than the critical value, reject H_0.

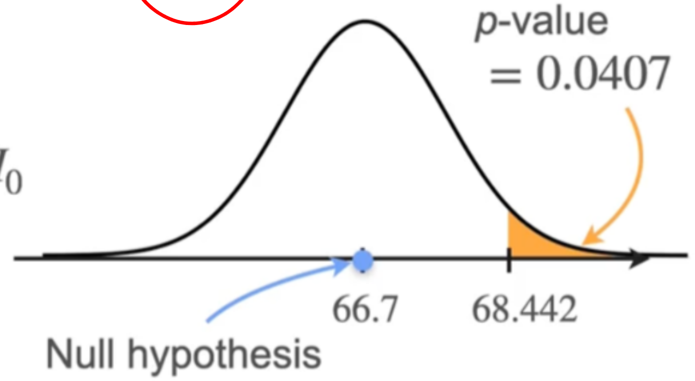
Both methods should lead to the same conclusion:
Right-tailed test: Reject H_0 if the statistic is greater than the critical value.
Left-tailed test: Reject H_0 if the statistic is less than the critical value.
Two-tailed test: Reject H_0 if the statistic is either greater than the upper critical value or less than the lower critical value.

Critical values allow you to predefine the decision rule, making it easier to calculate Type 2 error probabilities.

P-Values and Critical Values

3

If p-value $< \alpha$ ⇨ Reject H_0



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

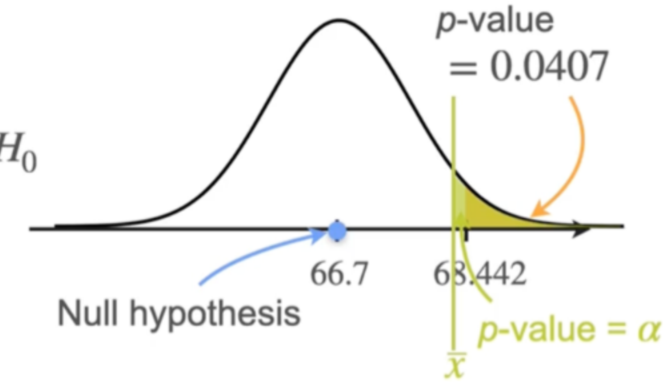
This is a sample that has the p-value of exactly Alpha.

DeepLearning.AI

4

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 $= \alpha$

Critical values

This is called a critical value.

DeepLearning.AI

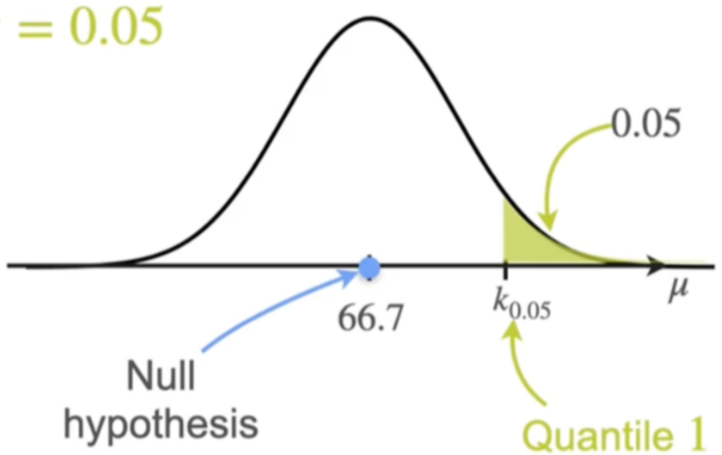
Computing Critical Values

5

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

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

$\alpha = 0.05$



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

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

of three divided by root 10.

DeepLearning.AI

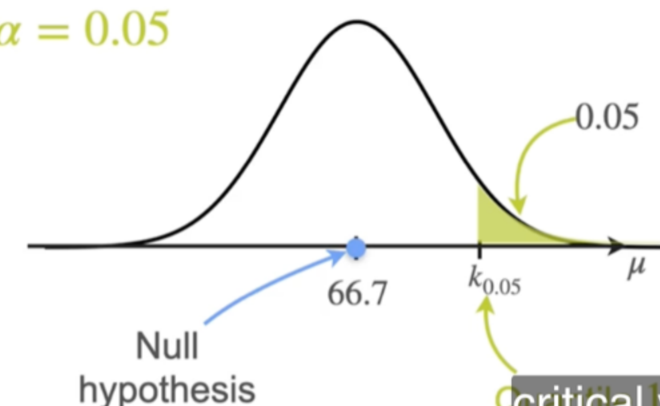
6

Computing Critical Values

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

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

$\alpha = 0.05$



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

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

$k_{0.05} = 68.26$

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

critical values is that you can define

DeepLearning.AI

This explanation covers two important concepts in hypothesis testing: **p-value** and **critical value**. Both are tools used to make decisions about whether to reject the null hypothesis (H_0) or not, but they work in slightly different ways. Let's break it down with simpler examples and key ideas:

1

1. P-value Method:

- **What it is:** The p-value represents the probability of obtaining a sample statistic at least as extreme as the one observed, given that the null hypothesis (H_0) is true.
- **How it works:**
  - If the p-value is **smaller than or equal to** the significance level (α, like 0.05), you **reject** H_0.
  - If the p-value is **greater than** α, you **fail to reject** H_0.

In your example, you calculated the **observed statistic** (the sample mean) and got a **p-value**. Based on the p-value, you decide whether or not to reject the null hypothesis.

2. Critical Value Method:

The **critical value** is like a "cutoff" that helps you decide whether to reject the null hypothesis. It depends on the significance level (α) you choose and the type of test you're performing (right-tailed, left-tailed, or two-tailed).

- **What it is:** The critical value is a point (or points) on the distribution that separates the **rejection region** from the **non-rejection region**. If your test statistic is more extreme than the critical value, you reject H_0.

Example of Critical Values:

1. **Right-tailed test** (testing if the sample mean is significantly greater than the population mean):

rejection region from the **non-rejection region**. If your test statistic is more extreme than the critical value, you reject H_0.

2

Example of Critical Values:

1. **Right-tailed test** (testing if the sample mean is significantly greater than the population mean):

   - If you're using an $\alpha = 0.05$ and want to find the critical value ($K_{0.05}$), you look for the value on the distribution that leaves **5% of the area** to the right.
   - In your case, that critical value is **68.26** (which means if your observed sample mean is **greater than 68.26**, you reject H_0).
2. **Left-tailed test** (testing if the sample mean is significantly smaller than the population mean):
   - You find the value where **5% of the area** is to the left.
   - The decision rule is: reject H_0 if the observed statistic is smaller than the critical value.
3. **Two-tailed test** (testing if the sample mean is **different** from the population mean):
   - Here, the area of α is divided between both sides (left and right), so you use two critical values.
   - If the observed statistic is **either greater than the upper critical value** or **smaller than the lower critical value**, you reject H_0.

Critical Value Example:

For your **right-tailed test** with $\alpha = 0.05$, the critical value was **68.26**. This means:

- **Reject** H_0 if the observed sample mean is greater than 68.26.
- If the sample mean is **less than or equal to** 68.26, you **fail to reject** H_0.

3

- If the sample mean is **less than or equal to** 68.26, you **fail to reject** H_0.

If you change α to 0.01, then the critical value moves to **68.91** (because with a smaller α, you require stronger evidence to reject H_0).

In your case, the observed sample mean was **68.442**, which is **greater than 68.26** but **less than 68.91**, so you **reject** H_0 at $\alpha = 0.05$ but **fail to reject** H_0 at $\alpha = 0.01$.

Why Critical Values are Useful:

- Critical values allow you to **predefine your decision rule** without having to calculate a p-value each time.
- This is useful for determining **Type 2 error probabilities** (failing to reject H_0 when H_1 is true) and helps in designing experiments.

Comparison between p-value and Critical Value:

Both methods should lead you to the **same conclusion**.

- If you **reject** H_0 based on the p-value, you should also reject H_0 based on the critical value.
- Similarly, if the p-value suggests **failure to reject** H_0, the critical value method will also lead you to **fail to reject** H_0.

Summary of Key Points:

- **P-value**: Measures how extreme the data is under H_0. If it's smaller than α, reject H_0.
- **Critical value**: Predefined cutoff based on α. If your statistic is more extreme than the critical value, reject H_0.
- Both methods lead to the **same conclusion** if used correctly.

4

In hypothesis testing, **p-values** and **critical values** are used to make decisions:

1. **P-value**: Measures the probability of observing a sample statistic as extreme as the one found, assuming the null hypothesis is true. If the p-value is smaller than the significance level (α), you reject the null hypothesis (H_0).
2. **Critical value**: The threshold that defines the rejection region. It corresponds to a value that leaves a specific area (like α) to the right (for a right-tailed test) or left (for a left-tailed test). This value depends on α. Anything more extreme than this value leads to rejecting H_0.

Example: Right-Tailed Test

- Null hypothesis: The mean height is 66.7.
- Alternative hypothesis: The mean height is greater than 66.7.
- Sample size: $n = 10$, standard deviation: $\sigma = 3$, $\alpha = 0.05$.
- **Critical value**: Find the value where 5% of the distribution lies to the right. This critical value is 68.26.

You can define your decision rule beforehand: **Reject H_0 if the sample mean is greater than 68.26**. With an observed sample mean of 68.442, you reject H_0 since it exceeds 68.26.

Changing α

If α changes to 0.01, the critical value increases (68.91), making it harder to reject H_0. In this case, the observed mean (68.442) is not enough to reject H_0.

Critical Value in Different Tests:

- **Right-tailed test**: Reject H_0 if the statistic is greater than the critical value.
- **Left-tailed test**: Reject H_0 if the statistic is smaller than the critical value.
- **Two-tailed test**: Reject H_0 if the statistic is greater than the upper critical value or smaller than the lower critical value.

5

Example: Right-Tailed Test

- Null hypothesis: The mean height is 66.7.
- Alternative hypothesis: The mean height is greater than 66.7.
- Sample size: $n = 10$, standard deviation: $\sigma = 3$, $\alpha = 0.05$.
- **Critical value**: Find the value where 5% of the distribution lies to the right. This critical value is 68.26.

You can define your decision rule beforehand: **Reject H_0 if the sample mean is greater than 68.26**. With an observed sample mean of 68.442, you reject H_0 since it exceeds 68.26.

Changing α

If α changes to 0.01, the critical value increases (68.91), making it harder to reject H_0. In this case, the observed mean (68.442) is not enough to reject H_0.

Critical Value in Different Tests:

- **Right-tailed test**: Reject H_0 if the statistic is greater than the critical value.
- **Left-tailed test**: Reject H_0 if the statistic is smaller than the critical value.
- **Two-tailed test**: Reject H_0 if the statistic is greater than the upper critical value or smaller than the lower critical value.

Key Points:

- You can define the critical value before collecting data, making it possible to compute **Type 2 error probabilities** (failing to reject H_0 when it's false).
- Both the p-value and critical value methods should give the same conclusion.

6

Critical values allow you to predefine the decision rule, making it easier to calculate Type 2 error probabilities.

Sources: Coursera Probability & Statistics for Machine Learning and Data Science.