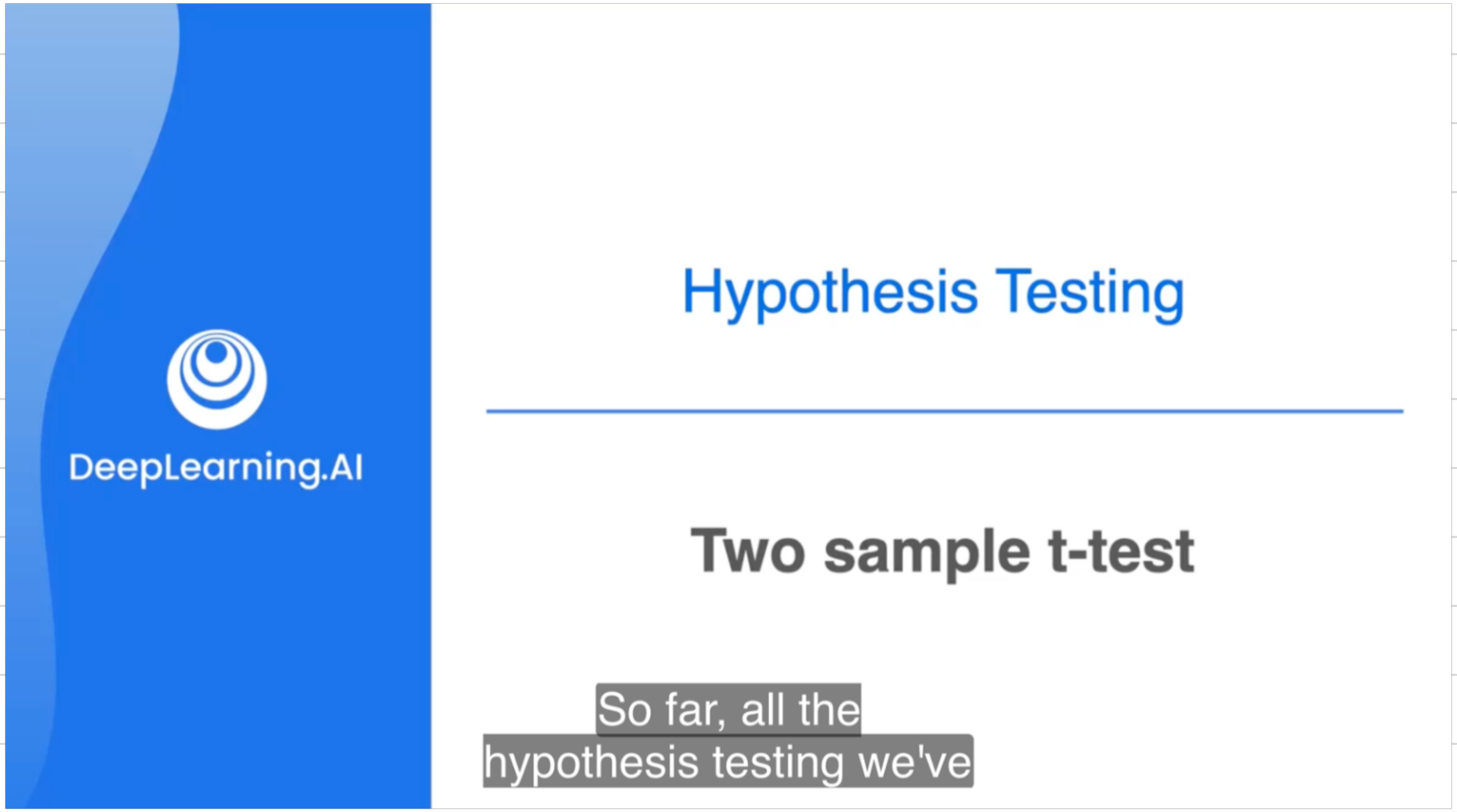
Day-69, Feb 07, 2025 (Magh 25, 2081)

1. Hypothesis testing: t-test (two sample t-test)
2. Paired t-test & two sample t-test
3. A/B testing: Machine Learning Application

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



Applications of Hypothesis Testing in Machine Learning:

1. Model Comparison
2. Feature Selection
3. Evaluating Model Accuracy
4. Testing Assumptions

Commonly used tests are

- t-test
- z-test
- ANOVA
- Chi-square

Problem Recap:

- Compare heights of 18-year-olds from the US and Argentina.
- US sample: $n_x = 10, \bar{x} = 68.442$ inches, $s_x = 3.113$ inches
- Argentina sample: $n_y = 9, \bar{y} = 65.949$ inches, $s_y = 3.106$ inches

---

Key Assumptions:

1. **Independent samples:** No measurement belongs to both groups.
2. **Independent observations:** Heights within each group are independent.
3. **Normality:** Heights in both populations follow Gaussian distributions.

Hypotheses for the Right-Tailed Test:

- **Null hypothesis (H_0):** $\mu_{US} - \mu_{Argentina} = 0$
- **Alternative hypothesis (H_a):** $\mu_{US} - \mu_{Argentina} > 0$

Test Statistic:

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Using the given values:

$$T = \frac{}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Using the given values:

$$T_{obs} = 1.7459$$

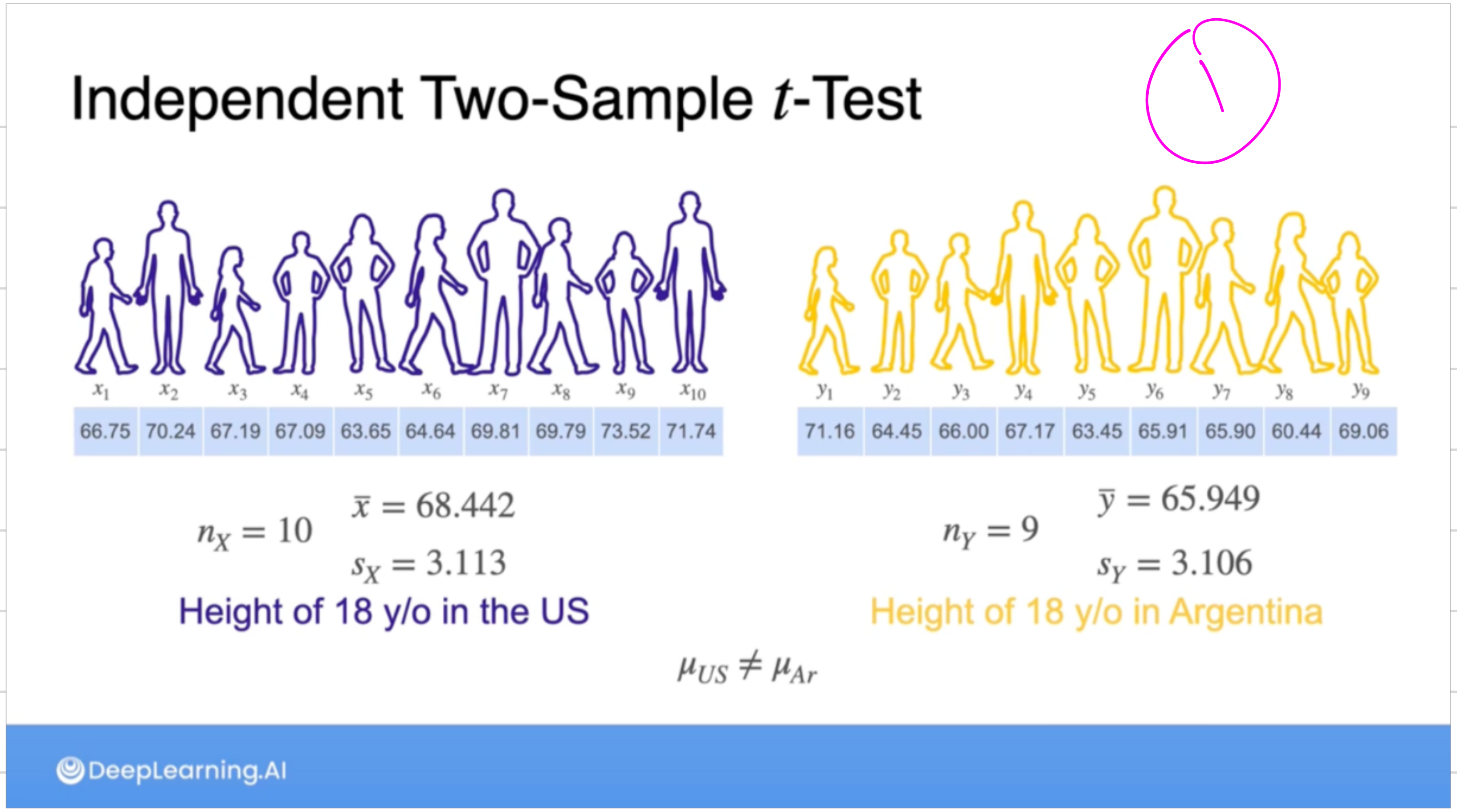
Degrees of freedom (computed via software): 16.8

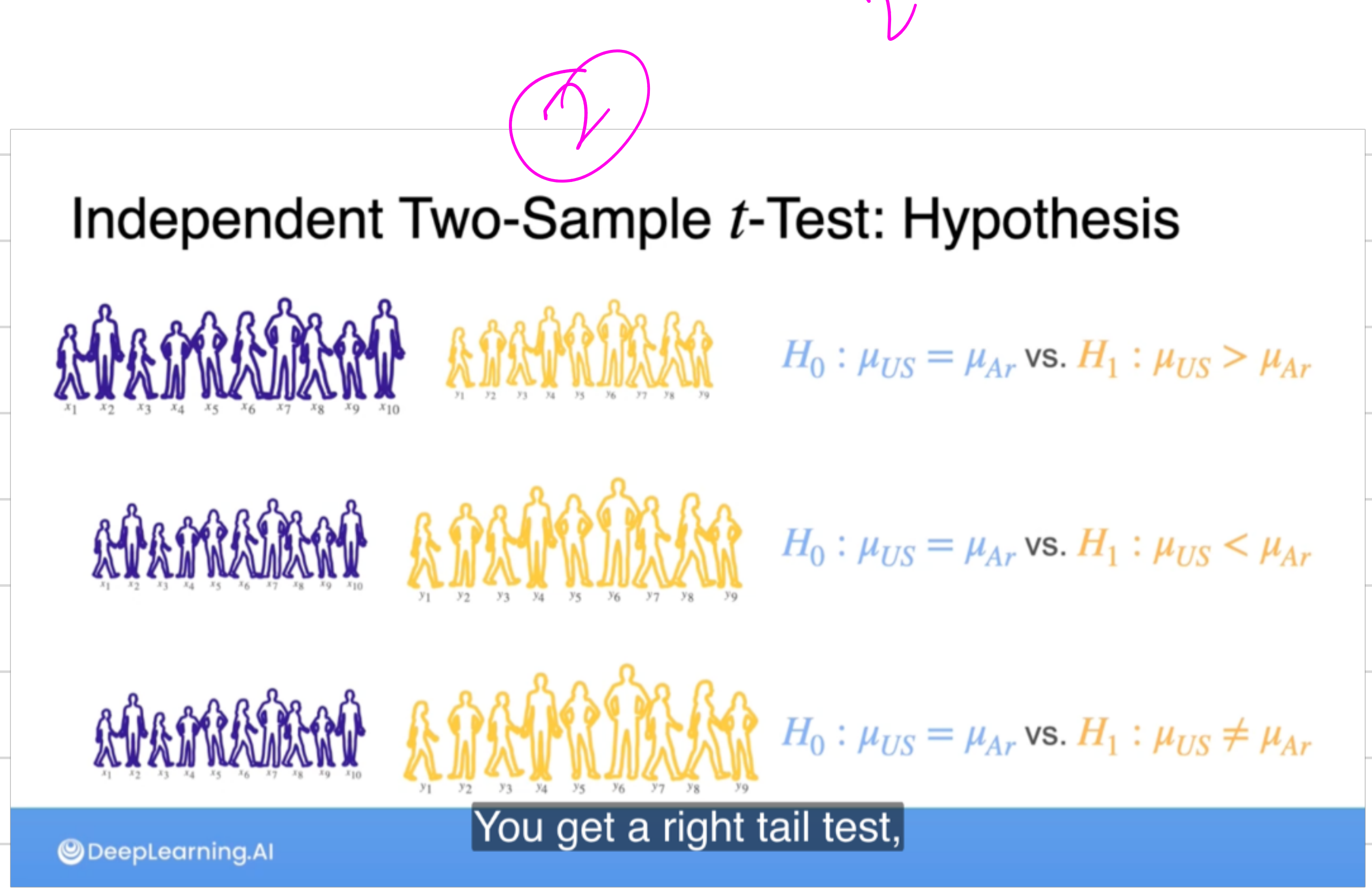
Right-Tailed Test Result:

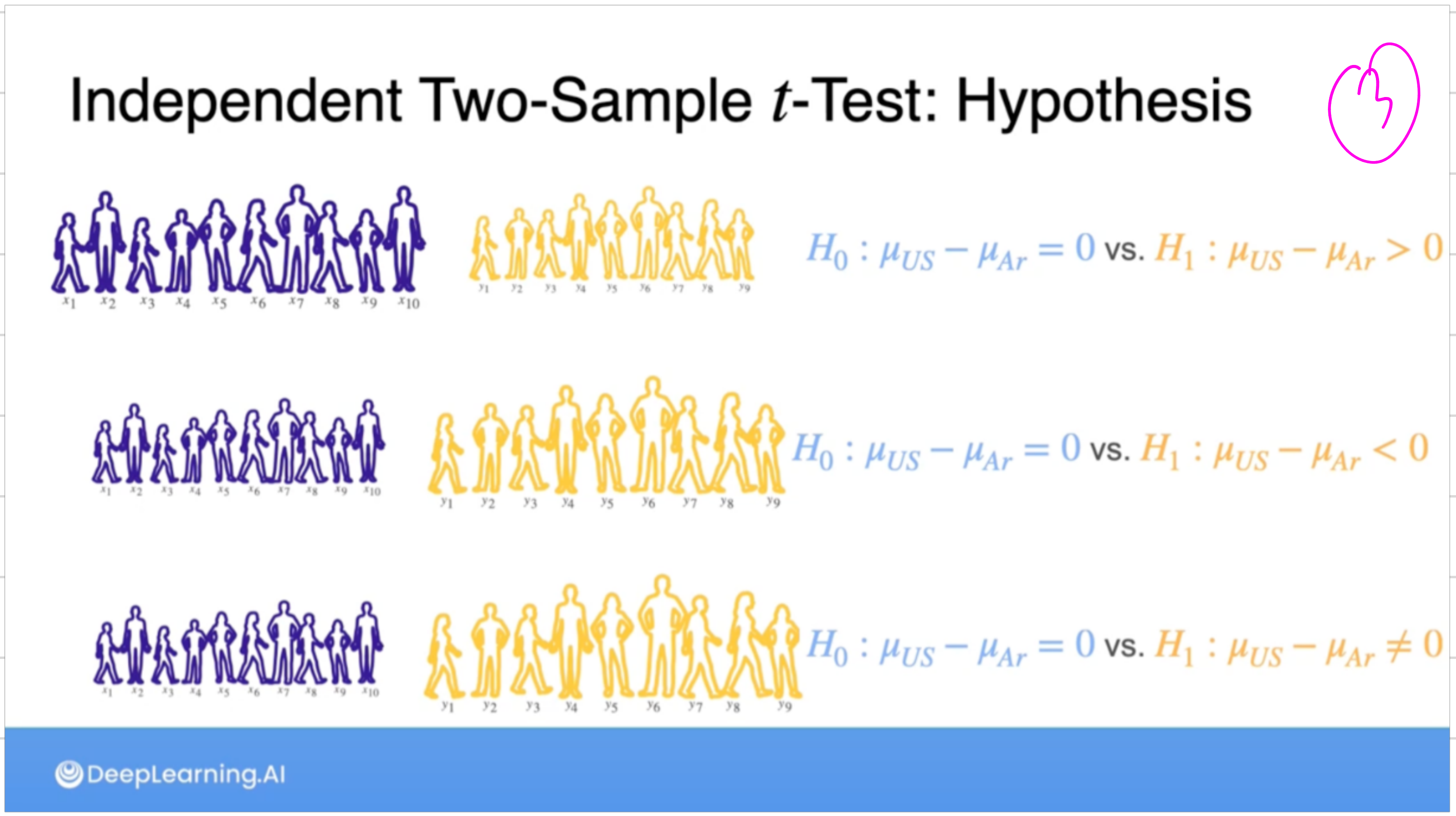
- **P-value:** 0.0495
- **Decision:** Since $P = 0.0495 < \alpha = 0.05$, reject H_0.
- **Conclusion:** The US population mean height is significantly greater than that of Argentina.

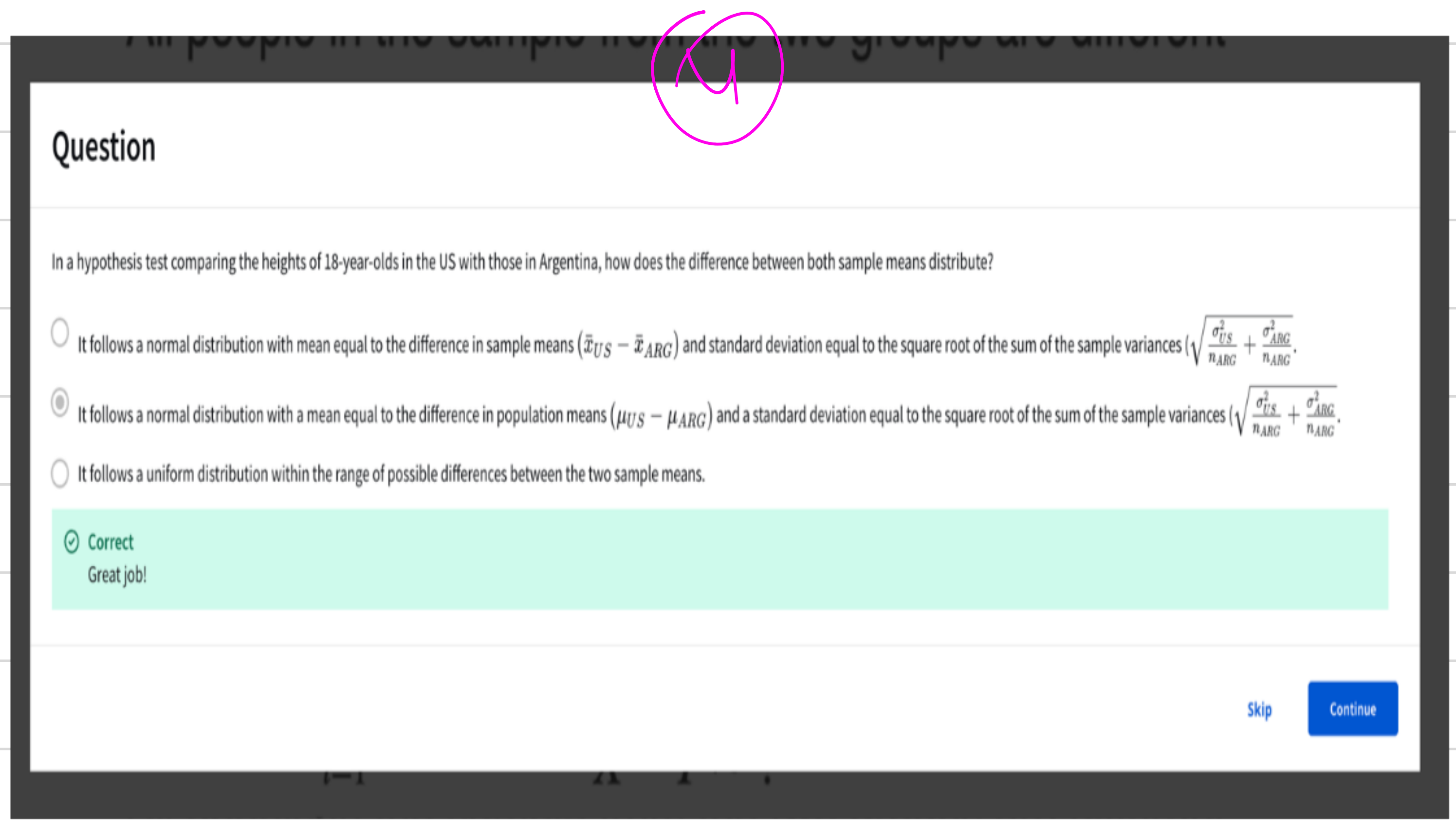
Two-Tailed Test Result:

- **Hypotheses:**
  - $H_0 : \mu_{US} - \mu_{Argentina} = 0$
  - $H_a : \mu_{US} \neq \mu_{Argentina}$
- **P-value:** 0.0991
- **Decision:** Since $P = 0.0991 > \alpha = 0.05$, fail to reject H_0.
- **Conclusion:** There's not enough evidence to say the population means are different.


1
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
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
$n_X = 10$
$\bar{x} = 68.442$
$s_X = 3.113$
Height of 18 y/o in the US
$n_Y = 9$
$\bar{y} = 65.949$
$s_Y = 3.106$
Height of 18 y/o in Argentina
$\mu_{US} \neq \mu_{Ar}$
DeepLearning.AI


2
Independent Two-Sample t-Test: Hypothesis
$H_0 : \mu_{US} = \mu_{Ar}$ vs. $H_1 : \mu_{US} > \mu_{Ar}$
$H_0 : \mu_{US} = \mu_{Ar}$ vs. $H_1 : \mu_{US} < \mu_{Ar}$
$H_0 : \mu_{US} = \mu_{Ar}$ vs. $H_1 : \mu_{US} \neq \mu_{Ar}$
You get a right tail test,
DeepLearning.AI


3
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. $H_1 : \mu_{US} - \mu_{Ar} \neq 0$
DeepLearning.AI


4
Question
In a hypothesis test comparing the heights of 18-year-olds in the US with those in Argentina, how does the difference between both sample means distribute?
It follows a normal distribution with mean equal to the difference in sample means $(\bar{x}_{US} - \bar{x}_{ARG})$ and standard deviation equal to the square root of the sum of the sample variances $(\sqrt{\frac{\sigma_{US}^2}{n_{ARG}} + \frac{\sigma_{ARG}^2}{n_{ARG}}}$.
It follows a normal distribution with a mean equal to the difference in population means $(\mu_{US} - \mu_{ARG})$ and a standard deviation equal to the square root of the sum of the sample variances $(\sqrt{\frac{\sigma_{US}^2}{n_{ARG}} + \frac{\sigma_{ARG}^2}{n_{ARG}}}$.
It follows a uniform distribution within the range of possible differences between the two sample means.
Correct
Great job!
Skip
Continue

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

5

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

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

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

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

then it will also be a Gaussian,

DeepLearning.AI

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

6

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

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

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

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

the square root of the
variances of each sample mean.

DeepLearning.AI

Independent Two-Sample t-Test: Statistic

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

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

You don't know σ_{US}, σ_{Ar} 😱

7

from any of the two groups,

DeepLearning.AI

Independent Two-Sample t-Test: Statistic

8

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

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

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

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

a Gaussian population
with an unknown Sigma.

DeepLearning.AI

Independent Two-Sample t-Test: Statistic

9

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

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

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

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

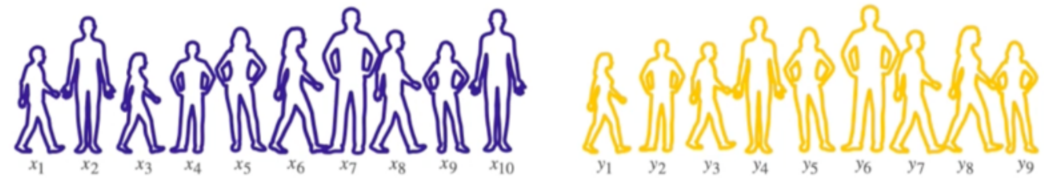
$$\text{Degrees of freedom} = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_y}\right)^2}{\frac{\left(\frac{s_X^2}{n_X}\right)^2}{n_X - 1} + \frac{\left(\frac{s_Y^2}{n_Y}\right)^2}{n_Y - 1}}$$

the degrees of freedom
is very cumbersome,

DeepLearning.AI

Independent Two-Sample t-Test

10



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

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

$$\text{Degrees of freedom} = \frac{\overbrace{\left(\frac{3.113^2}{10} + \frac{3.106^2}{9}\right)^2}^{16.8}}{\frac{\left(\frac{3.113^2}{10}\right)^2}{10-1} + \frac{\left(\frac{3.106^2}{9}\right)^2}{9-1}}$$

s_x and x_y by the
observed values,

DeepLearning.AI

Independent Two-Sample t-Test: Statistic

11

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

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

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

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

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

16.8 Now after all that math,

DeepLearning.AI

Independent Two-Sample t-Test: Right Tailed Test

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

$\alpha = 0.05$

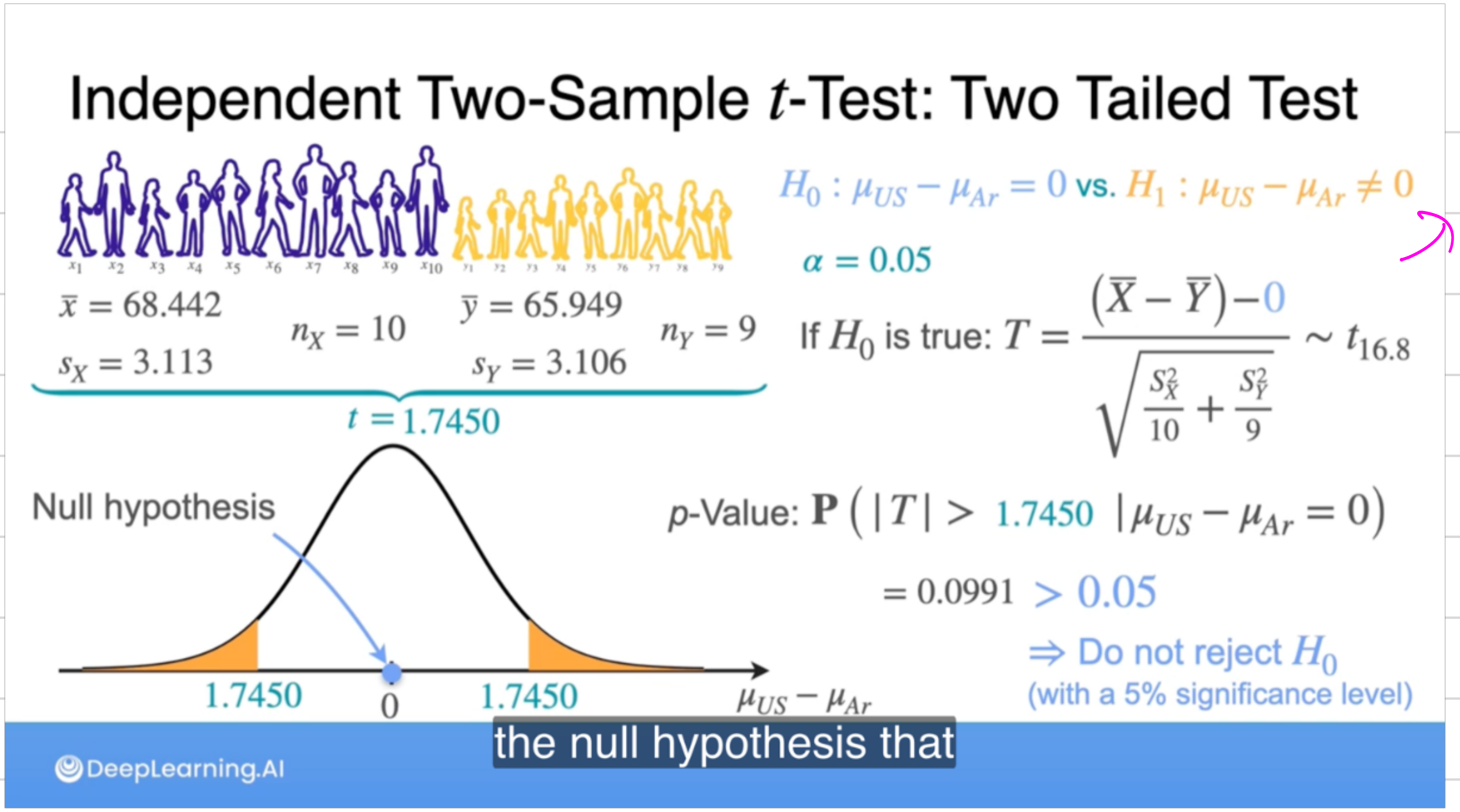
$\bar{x} = 68.442$ $\quad s_X = 3.113$ $\quad n_X = 10$

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

12

From last slide you know
that if H_0 is true,

DeepLearning.AI



13 **Application of Statistics & Probability**

1. Data Distribution
2. Model Uncertainty
3. Bayesian Inference
4. Dimensionality Reduction

Algorithms that uses Statistics and Reliability

1. Linear Regression
2. Logistic Regression
3. Time-Series forecasting (ARIMA)
4. Expectation-Maximization.

Hypothesis Testing

Paired t-test

two test are Paired if basketball team are split into half

Both are Paired t-test,

This explanation dives into the **Paired Sample t-Test**, where paired observations are made on the same subjects before and after a treatment (like measuring weights before and after a training program). Let's break it down step by step and add a simple Python code example.

Concept Recap

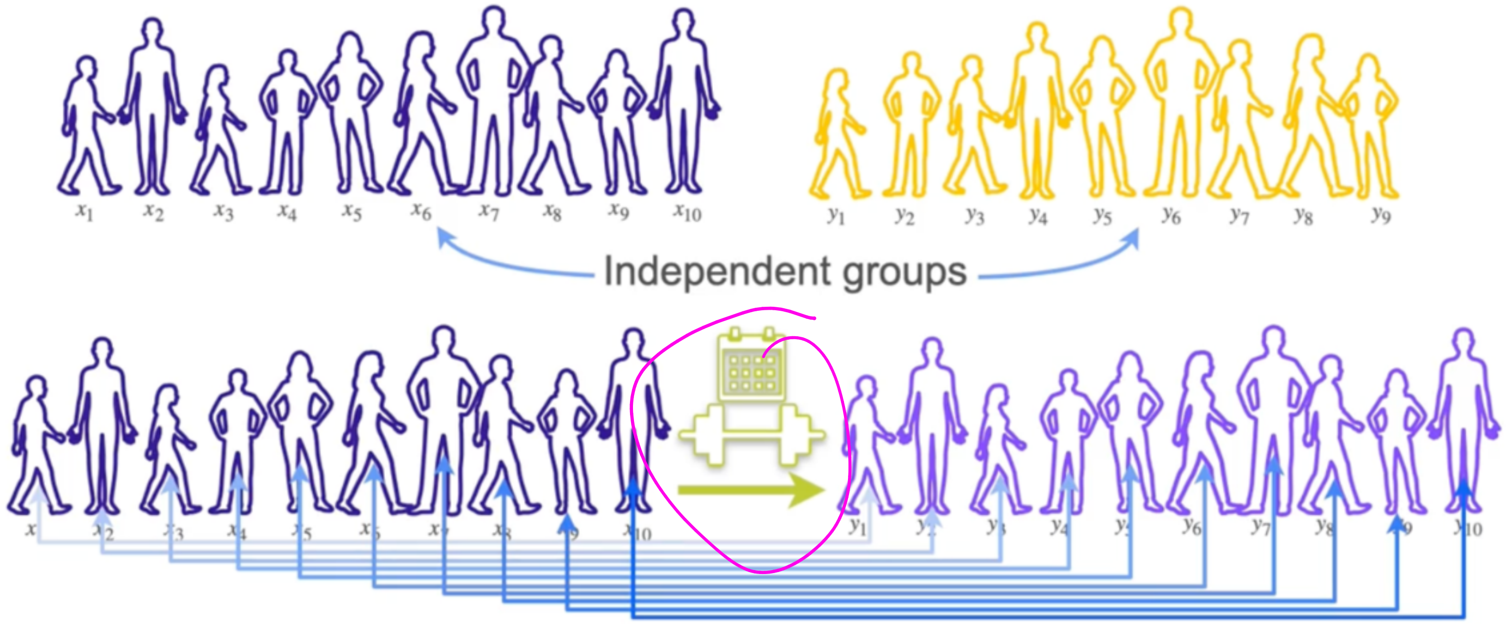
1. **Purpose:**
   - To determine if there is a significant difference between two related (paired) groups.
2. **How It Works:**
   - Compute the differences for each pair, $D_i = X_i - Y_i$.
   - Calculate the mean of these differences ($\bar{D}$) and the sample standard deviation (s_D).
   - Use the t-statistic:

$$t = \frac{\bar{D}}{s_D/\sqrt{n}}$$

   - Compare the t-statistic against the critical t-value or compute the p-value.
3. **Hypotheses:**
   - $H_0 : \mu_D = 0$ (No difference in means)
   - $H_a : \mu_D > 0$ (Mean difference is positive, i.e., the program is effective)

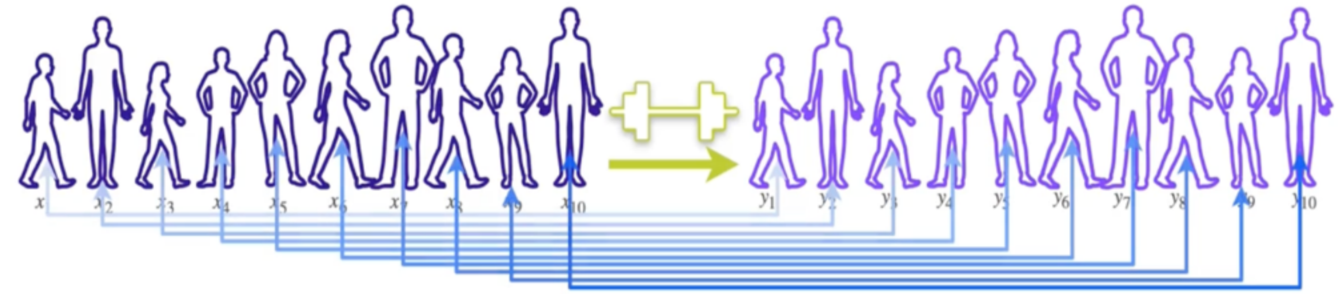
Paired t-Test and Two-Sample t-Test

1



Paired t-Test: Statistic

2



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

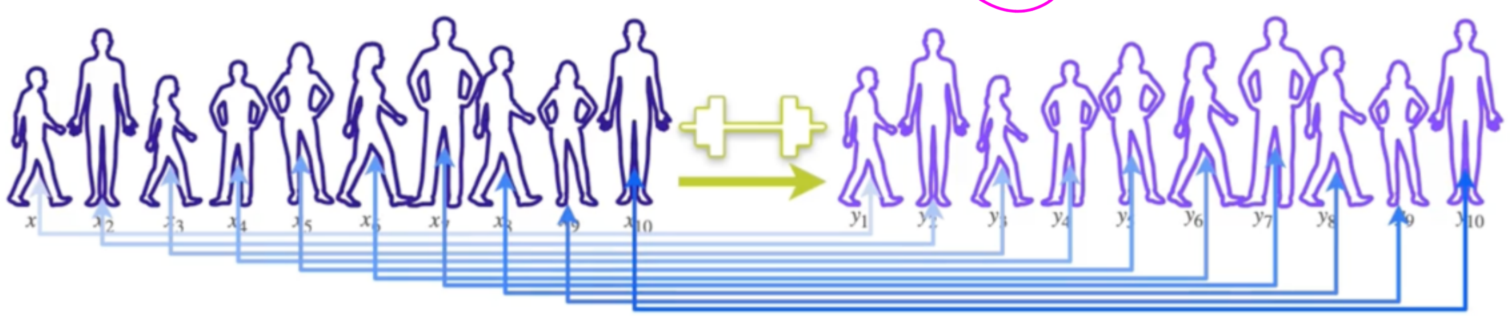
$$\frac{(X_1 - Y_1)+(X_2 - Y_2)+(X_3 - Y_3)+(X_4 - Y_4)+(X_5 - Y_5)+(X_6 - Y_6)+(X_7 - Y_7)+(X_8 - Y_8)+(X_9 - Y_9)+(X_{10} - Y_{10})}{10}$$

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

Sum Because of Same Group Combinations.

Paired t-Test: Statistic

3



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

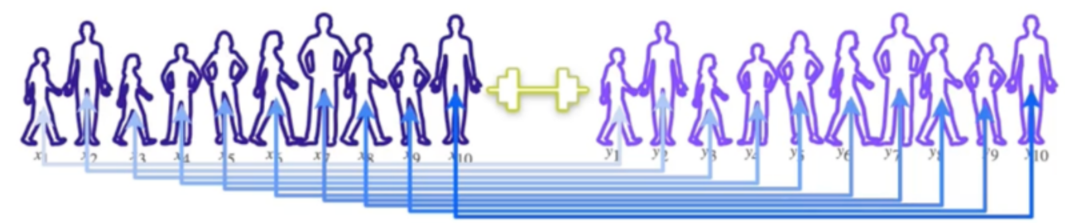
$$\overline{D} = \frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10}$$

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

4



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

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

$D_i = X_i - Y_i$

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

$\dfrac{\overline{D} - \mu_D}{\sigma_D/\sqrt{10}} \sim \mathcal{N}(0,1^2)$ But σ_D is unknown $\Rightarrow \sigma_D \rightarrow S_D = \sqrt{\dfrac{\sum_{i=1}^{10}(D_i - \overline{D})^2}{10-1}} \Rightarrow T = \dfrac{\overline{D} - \mu_D}{S_D/\sqrt{10}}$

Hypothesis Setup

- **Null Hypothesis (H_0)**: The mean difference between weights is zero ($\mu_D = 0$).
  This implies the training program has no effect.
- **Alternative Hypothesis (H_a)**: The mean difference is positive ($\mu_D > 0$).
  This implies the training program is effective in reducing weight.

5

Test Statistic Calculation

1. Compute the **mean of the differences ($\bar{D}$)**:

$$\bar{D} = \frac{\sum D_i}{n}$$

   In this case, the sample mean difference is **1.09**.

2. Compute the **sample standard deviation of the differences (s_D)**, which is **1.485**.
3. Compute the **t-statistic** using the formula:

$$t = \frac{\bar{D}}{s_D/\sqrt{n}}$$

   where n is the number of paired samples (10 in this case).

4. Plugging in the values:

$$t = \frac{1.09}{1.485/\sqrt{10}} \approx 2.321$$

$1.485/\sqrt{10}$

6

---

p-Value and Decision

- The p-value is the probability of getting a t-statistic as extreme as 2.321 under the null hypothesis.
- For a **right-tailed test**, this corresponds to a p-value of **0.0227**.
- Since **0.0227 < 0.05 (significance level)**, you **reject the null hypothesis**.

Conclusion

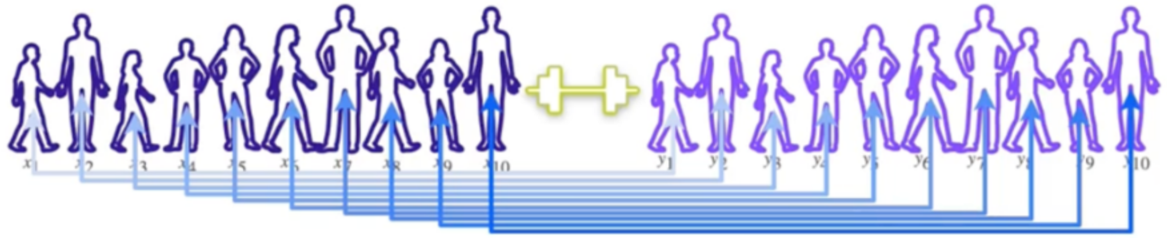
The decision to reject the null hypothesis implies that the training program is effective for weight loss on average.

Key Insight

- By transforming the problem into studying the differences (D_i) instead of comparing two separate groups, the paired sample t-test effectively reduces to a **single sample t-test** on the differences.

7

Paired t-Test: Statistic



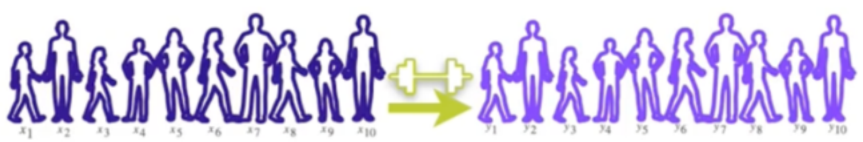
Now you're interested in the difference between pair of samples

$$\overline{D} = \frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10} \qquad D_i = X_i - Y_i \qquad D_i \overset{i.i.d}{\sim} \mathcal{N}(\mu_D, \sigma_D^2)$$

$$\frac{\overline{D} - \mu_D}{\sigma_D / \sqrt{10}} \sim \mathcal{N}(0, 1^2)$$ But σ_D is unknown $\Rightarrow \sigma_D \to S_D = \sqrt{\frac{\sum_{i=1}^{10} (D_i - \overline{D})^2}{10 - 1}} \Rightarrow T = \frac{\overline{D} - \mu_D}{S_D / \sqrt{10}}$

DeepLearning.AI

Paired Two-Sample t-Test: Right Tailed Test



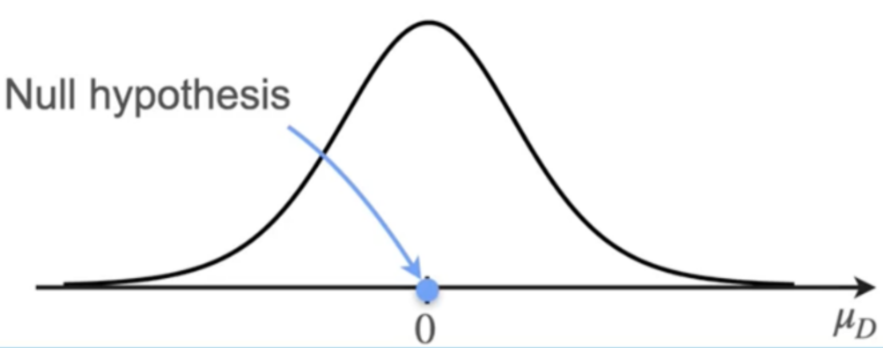
$\bar{d} = 1.09$ $\quad s_D = 1.485$ $\quad n = 10$

$t = 2.321$

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

$\alpha = 0.05$

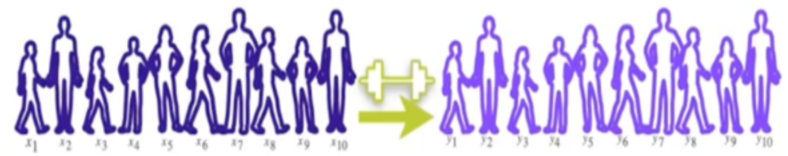
If H_0 is true: $T = \frac{\overline{D} - 0}{S_D / \sqrt{10}} \sim t_{10-1}$



8

9

Paired Two-Sample t-Test: Right Tailed Test



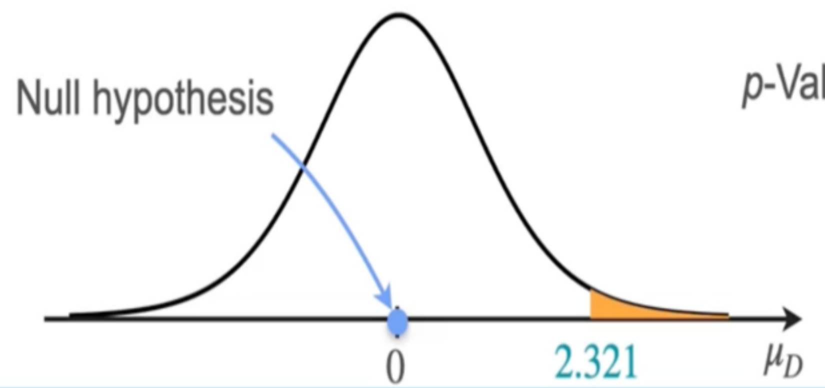
$\bar{d} = 1.09$ $\quad s_D = 1.485$ $\quad n = 10$

$t = 2.321$

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

$\alpha = 0.05$

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



p-Value: $\mathbf{P}\left(T > 2.321 \mid \mu_D = 0\right)$

$= 0.0227 < 0.05$

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

DeepLearning.AI

This means there is a difference.

This content explains A/B testing, a practical application of two-sample hypothesis testing used to compare two website designs (A and B). Here's a concise summary:

1. **Scenario**: A company tests if design B, with a different "Buy Now" button placement, leads to higher purchases compared to design A.
2. **Data Collection**:
   - Design A: 80 users, average purchase \$50, standard deviation \$10
   - Design B: 20 users, average purchase \$55, standard deviation \$15
3. **Hypothesis**:
   - Null (H_0): Mean purchase amounts are equal
   - Alternative (H_1): Mean for B is higher
4. **T-test Application**: Statistic = -1.414, p-value = 0.085 > 0.05 (alpha), so H_0 is not rejected.
5. **Conversion Rate Test:** Comparing conversion rates between designs using a binomial distribution test (20/80 for A, 8/20 for B).
   - Statistic = -1.336, p-value = 0.091, leading to no rejection of H_0.

Key Takeaway: A/B testing involves designing experiments, gathering data, and using hypothesis testing (like t-tests) to evaluate changes, with A/B methodology encompassing broader steps beyond statistical tests.

A/B Testing

ML Application

Got it! You've provided a comprehensive explanation of A-B testing and its relation to hypothesis testing, specifically two-sample hypothesis testing. Here's a summarized breakdown of the key points for clarity:

A-B Testing Overview

- **Scenario:** A website wants to test two different placements for a "Buy Now" button (Strategy A vs. Strategy B).
- **Goal:** Determine if switching to Strategy B leads to higher purchase amounts.
- **Method:** Randomly assign customers to Strategy A or B. Usually, a smaller group is tested on the new strategy.

Example 1: Purchase Amount Comparison

- Strategy A: 80 customers, average purchase of \$50, standard deviation of \$10
- Strategy B: 20 customers, average purchase of \$55, standard deviation of \$15
- **Hypotheses:**
  - H_0: Mean purchase of Strategy A = Mean purchase of Strategy B
  - H_A: Mean purchase of Strategy B > Strategy A
- **Alpha Value:** 0.05
- **Test Used:** Two-sample t-test assuming Gaussian distribution
- Degrees of Freedom: 23.38

Example 1: Purchase Amount Comparison

- Strategy A: 80 customers, average purchase of \$50, standard deviation of \$10
- Strategy B: 20 customers, average purchase of \$55, standard deviation of \$15
- **Hypotheses:**
  - H_0: Mean purchase of Strategy A = Mean purchase of Strategy B
  - H_A: Mean purchase of Strategy B > Strategy A
- **Alpha Value:** 0.05
- **Test Used:** Two-sample t-test assuming Gaussian distribution
- Degrees of Freedom: 23.38
- Observed Statistic: -1.414
- P-value: 0.085
- **Conclusion:** Do not reject H_0 since p-value > 0.05.

$-1.414.$

p-value > 0.05.

A-B Testing vs. T-tests

- A-B testing encompasses the entire experimental process:
  - Proposing variations
  - Random sample splitting
  - Outcome measurement
  - Decision-making with statistical tools
- The t-test is just one statistical tool used within A-B testing when comparing means.

Example 2: Conversion Rate Comparison

- Group A: 80 customers, 20 conversions
- Group B: 20 customers, 8 conversions
- **Hypotheses:**
  - H_0: Conversion rate for A = Conversion rate for B
  - H_A: Conversion rate for B > A
- **Alpha:** 0.05
- **Test Used:** Binomial distribution-based test
- Observed Statistic: -1.336
- P-value: 0.091
- **Conclusion:** Reject H_0 since p-value > 0.05, supporting that Strategy B may have a better conversion rate.

Reject H_0.
Accept H_A.
B

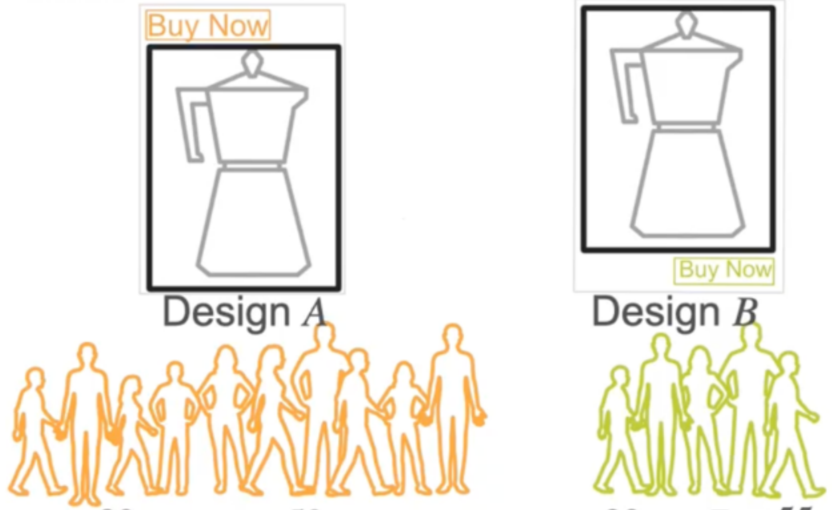
Strategy B is winner.

Key Concepts Covered

- Law of Large Numbers
- Central Limit Theorem
- Use of binomial distributions for conversion rates
- Estimation of p using combined observations
- Standardization of test statistics

1

A/B Testing: Purchase Amount

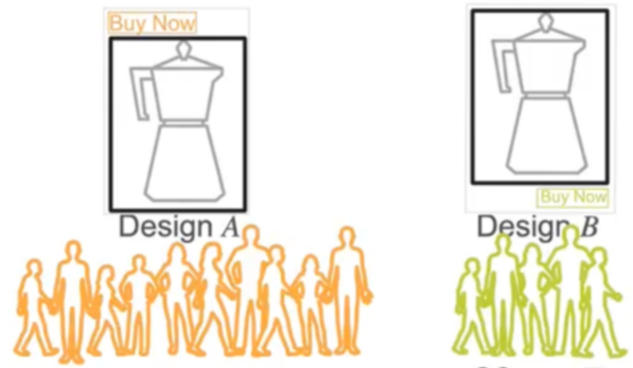


$n_A = 80$ $\quad \bar{x} = 50$ $\quad s_A = 10$

$n_B = 20$ $\quad \bar{y} = 55$ $\quad s_B = 15$

2

A/B Testing: Purchase Amount



$n_A = 80$ $\quad \bar{x} = 50$ $\quad s_A = 10$

$n_B = 20$ $\quad \bar{y} = 55$ $\quad s_B = 15$

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

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

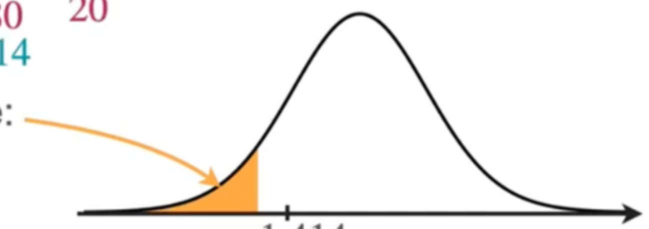
$\alpha = 0.05$

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

$t = \dfrac{(50 - 55) - 0}{\sqrt{\dfrac{10^2}{80} + \dfrac{15^2}{20}}}$

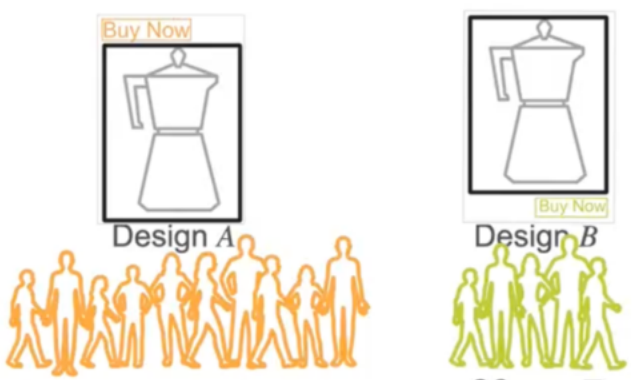
-1.414

p-Value:



3

A/B Testing: Purchase Amount



$n_A = 80$ $\quad \bar{x} = 50$ $\quad s_A = 10$

$n_B = 20$ $\quad \bar{y} = 55$ $\quad s_B = 15$

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

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

$\alpha = 0.05$

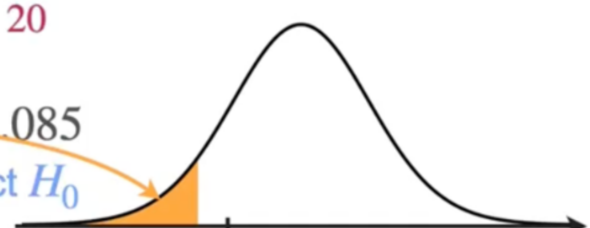
If H_0 is true: $T = \dfrac{(\overline{X} - \overline{Y}) - 0}{\sqrt{\dfrac{S_A^2}{n_A} + \dfrac{S_B^2}{n_B}}} \sim t_{23.38}$

$t = \dfrac{(50 - 55) - 0}{\sqrt{\dfrac{10^2}{80} + \dfrac{15^2}{20}}}$

-1.414

p-Value: 0.085

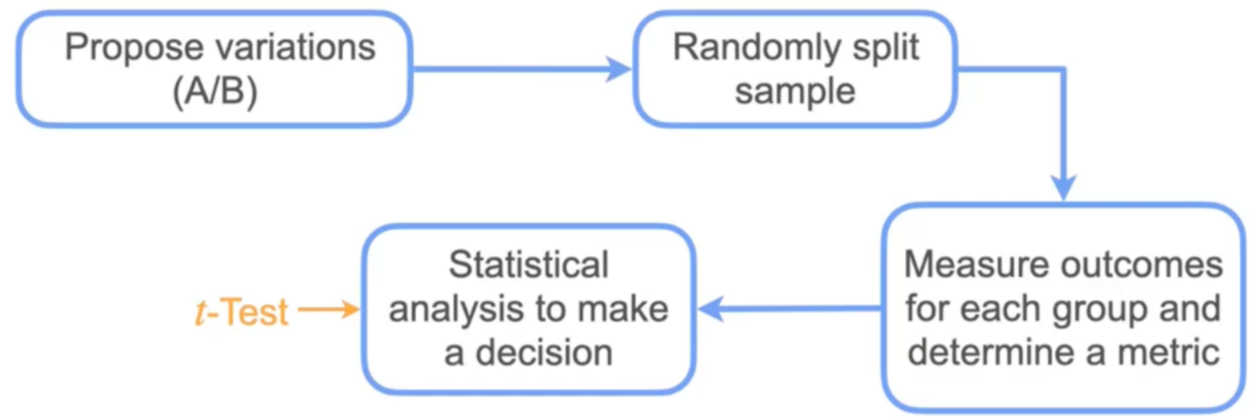
Don't reject H_0



4

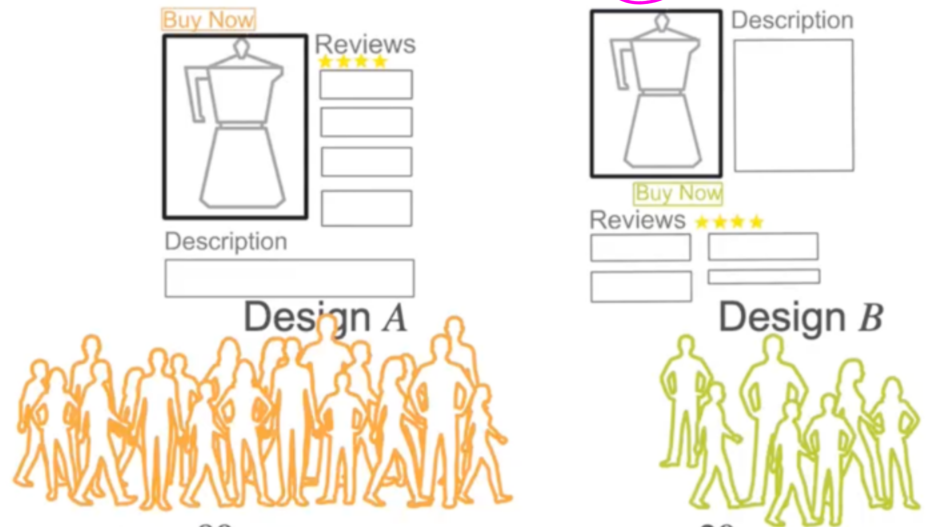
A/B Testing and *t*-Tests

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



5

A/B Testing: Conversion Rates

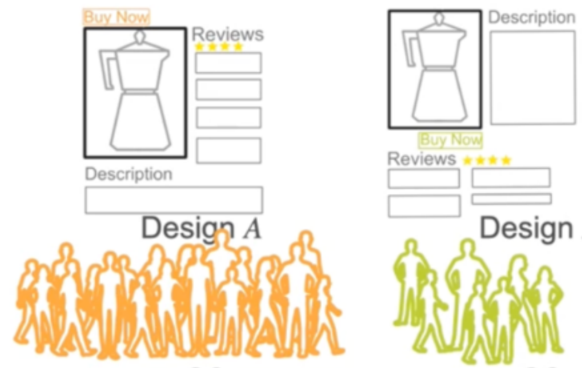


$n_A = 80$ $\qquad$ $n_B = 20$

DeepLearning.AI

6

A/B Testing: Conversion Rates



$n_A = 80$ $\qquad$ $n_B = 20$

$x = 20$ $\qquad$ $y = 8$

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

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

p_A = Conversion rate from Design A

p_B = Conversion rate from Design B

$\alpha = 0.05$

Now what is the
statistic to use in this case?

DeepLearning.AI

7

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.

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

These are good approximations
for pA and pB so x over n NA minus

DeepLearning.AI

8

A/B Testing: Conversion Rates

Statistic?

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

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

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

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

$$\Updownarrow$$

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

One very common way to express this
is by standardizing the distribution.

DeepLearning.AI

9

A/B Testing: Conversion Rates

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

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

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

1 over na plus 1 over mb can be written
as na plus mb divided by na times nb.

DeepLearning.AI

10

A/B Testing: Conversion Rates

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

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

But you don't know p

$\downarrow$

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

Test statistic

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

DeepLearning.AI

11

A/B Testing: Conversion Rates

$n_A = 80$ $\quad$ $n_B = 20$

$x = 20$ $\quad$ $y = 8$

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

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

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

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

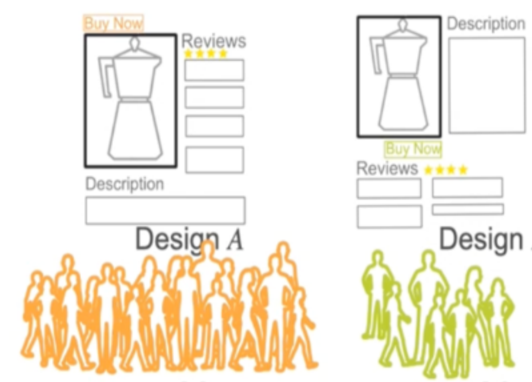
$$z = \frac{\left(\frac{20}{80}-\frac{8}{20}\right)-0}{\sqrt{(20+8)\left(1-\frac{20+8}{80+20}\right)}}\sqrt{80\,20}$$

$z = -1.336$

DeepLearning.AI

12

A/B Testing: Conversion Rates



$n_A = 80$ $\quad$ $n_B = 20$

$x = 20$ $\quad$ $y = 8$

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

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

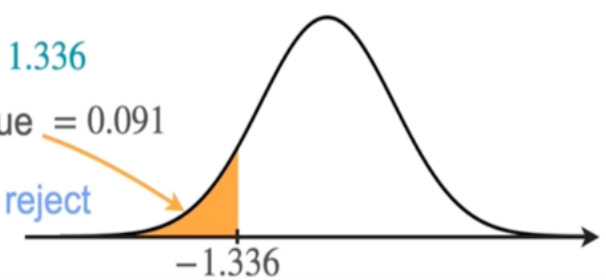
$\alpha = 0.05$ $\quad$ If H_0 is true $\Rightarrow p_A = p_B = p$

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

$z = -1.336$

p-value $= 0.091$

Do not reject H_0



DeepLearning.AI