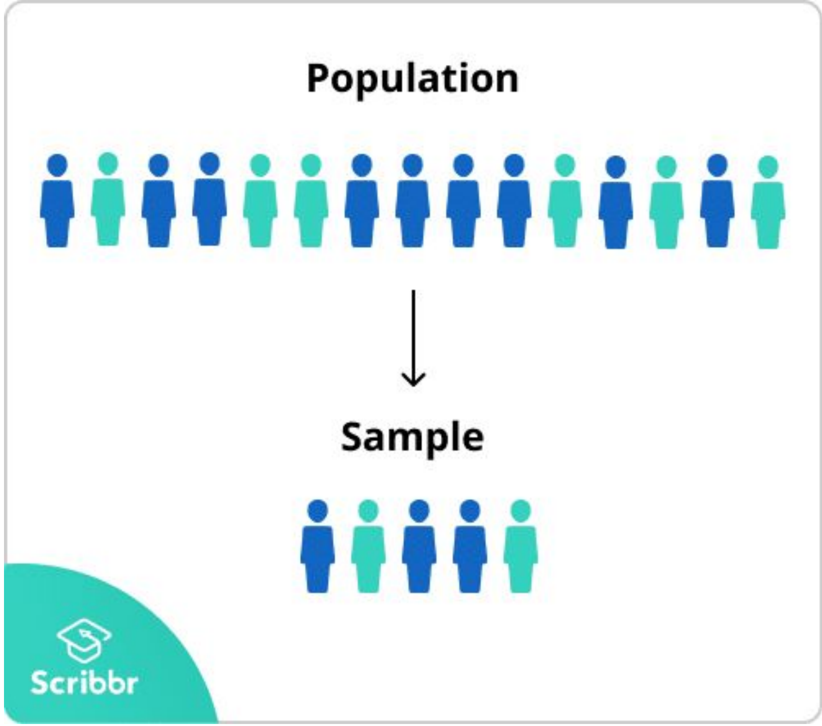
T-Test,
A Strong Way to Validate Your Experiment

Bahy Helmi Hartoyo Putra
ML Engineer @DOKU
github.com/bahyhelmihp/sesi-berbagi
Community of Practice Data Science

Adopted from:
Statistics for the Behavioral Science 9th ed.
by Frederick J. Gravetter, Larry B. Wallnau

Key Concepts



Population vs Sample

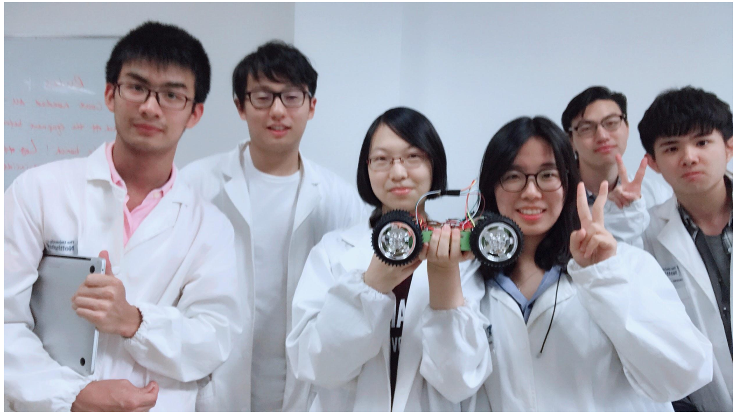
Key Concepts

Sample = [1, 2, 3, 3, 4, 4, 4, 5, 5, 6, 7]

Sample size (n) =
11
Sample mean (M) =
44/11 = **4**
Sum of Squares (SS) =
$(1-4)^2 + (2-4)^2 + \ldots = $ **30**
Sample variance (s^2) =
$(1-4)^2 + (2-4)^2 + \ldots / (11-1) = $ **3**

Sample Size,
Sample Mean,
Sample SS,
Sample Variance

Introduction

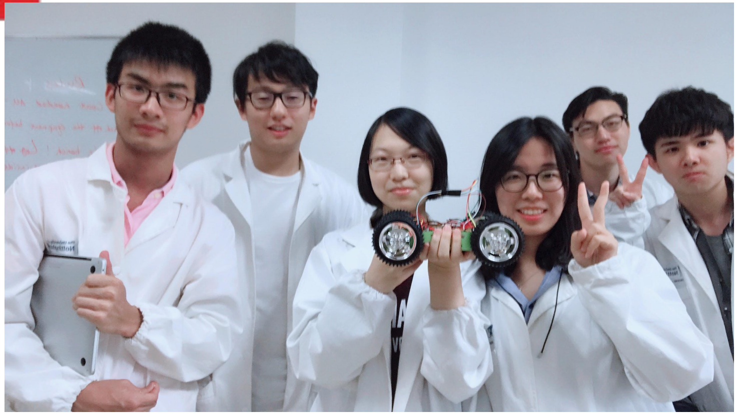


Group 1: Experiment-Based Learning

Group 2: Reading-Based Learning

Midterm Test Score

Introduction



Midterm Test Score Group #1:
Experiment-Based Learning
[70, 90, 80, 75, …], n = 15
M = 93

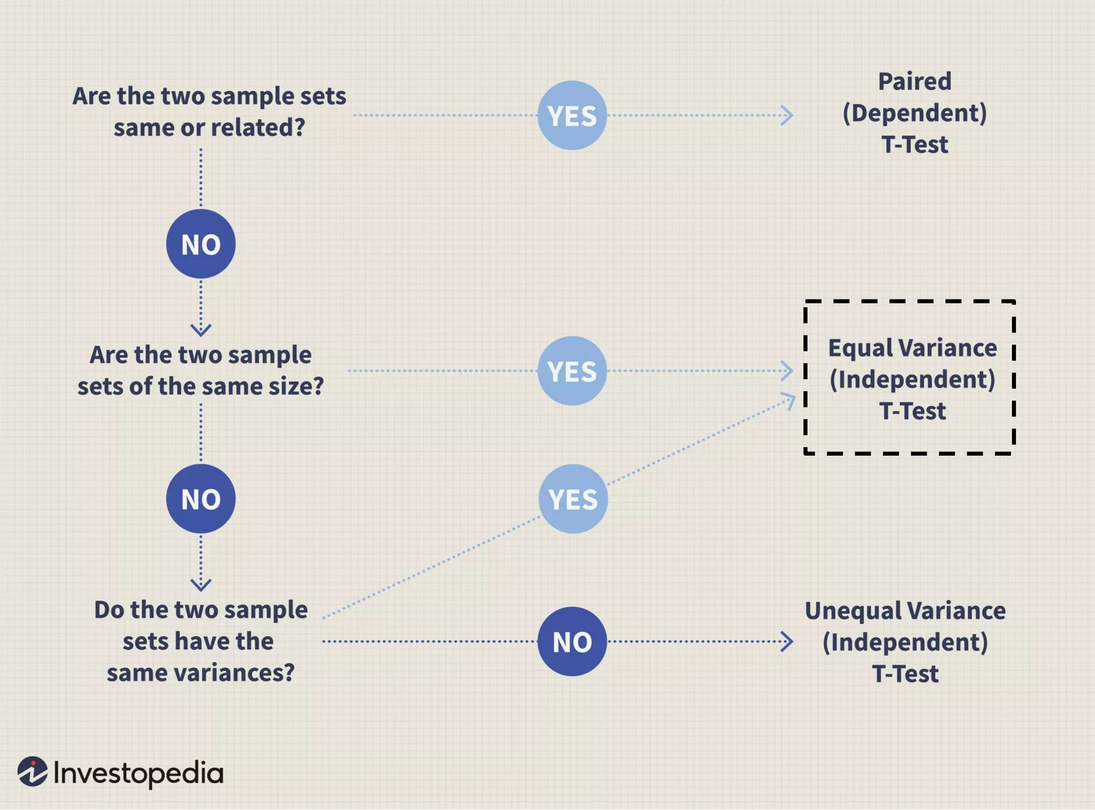
Midterm Test Score Group #2:
Reading-Based Learning
[75, 65, 80, 85, …], n = 15
M = 85

Experiment-Based shows higher average than Reading-Based.
Sure about this? Is this difference significant?

What is t-test?

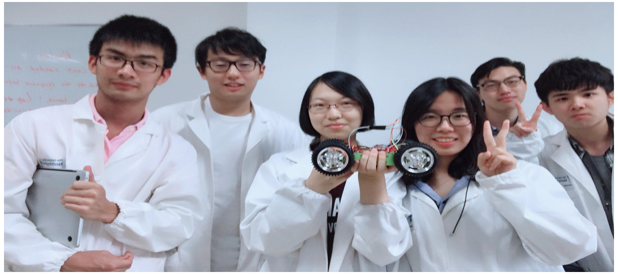
- A t-test is a **type of inferential statistic** used to determine if there is **a significant difference** between the means of two groups.
- T-test is used when:
  - **One or more** treatment(s) done to a population.
  - **Population variance** is **unknown**.
  - Sample size is **relatively small** (n <= 30)*.
- T-test exists for two kind of research design:
  - **Independent**-measures research design.
  - **Repeated**-measures research design.

Kind of T-Test



Two Independent Samples T-Test

Revisit Problem



Midterm Test Score Group #1:
[70, 90, 80, 75, …], n = 15
M = 93

Midterm Test Score Group #2:
[75, 65, 80, 85, …], n = 15
M = 85

Two possible explanations about the above case:

- It is possible that **there is really a difference between** the two treatment conditions so that the method used by **Group #1** produces better score than the method used by **Group #2**.
- It is possible that **there is no difference between** the the method used by **Group #1** and **Group #2**. The **mean difference obtained** in the experiment is simply the **result of sampling error.**

Steps on
Calculating Two Independent Samples T-Test

The Problem

Average High School Grade			
Experiment-based		**Reading-based**	
86	99	90	79
87	97	89	83
91	94	82	86
97	89	83	81
98	92	85	92
$n = 10$		$n = 10$	
$M = 93$		$M = 85$	
$SS = 200$		$SS = 160$	

Based on the data on the left, Is there a **significant difference** between the **two groups**? Use a **two-tailed test** with **α** = 0.01.

#1: Define Hypothesis

μ_1 = Population Mean of Experiment-based learning

μ_2 = Population Mean of Reading-based learning

$H_0: \mu_1 - \mu_2 = 0$ **or** $H_0: \mu_1 = \mu_2$

Null hypothesis:

- **No mean difference** between the population means.
- **Experiment-based** learning and **reading-based** learning comes from an exactly **same** population.

$H_1: \mu_1 - \mu_2 \neq 0$ **or** $H_0: \mu_1 \neq \mu_2$

Alternative hypothesis:

- **There is a mean difference** between the population means.
- **Experiment-based** learning and **reading-based** learning comes from a **different** population.

#2: Find the critical region

We set the alpha level for this experiment,
α = 0.01, two-tailed.

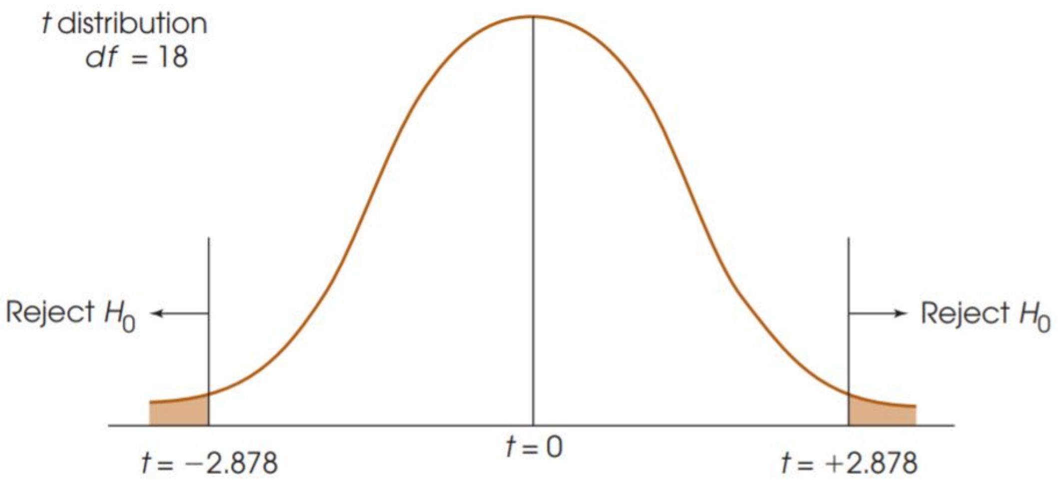
This is an independent-measures design. The t statistic for these data has degrees of freedom determined by

$$\begin{aligned} df &= df_1 + df_2 \\ &= (n_1 - 1) + (n_2 - 1) \\ &= 9 + 9 \\ &= 18 \end{aligned}$$

Given these informations, how we could find the **critical region**?

#2: Find the critical region (cont.)

The t distribution for **df = 18** is and **α = 0.01**, the critical region consists of the extreme **1% of the distribution** and has **boundaries** of **t = +2.878 and t = -2.878**.



#3: Find the t-statistic

$$t = \frac{\text{sample mean} - \text{population mean}}{\text{estimated standard error}} = \frac{M - \mu}{s_M}$$

$$t = \frac{\text{sample mean difference} - \text{population mean difference}}{\text{estimated standard error}} = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}}$$

#3: Find the t-statistic (cont.)

$$t = \frac{\text{sample mean difference} - \text{population mean difference}}{\text{estimated standard error}} = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\boxed{s_{(M_1 - M_2)}}}$$

$$s_{(M_1 - M_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{\boxed{s_p^2}}{n_2}}$$

$$\text{pooled variance} = s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

REMEMBER!

M = sample mean
s^2 = sample variance
s^2_p = pooled variance
n = sample size
SS = sum of squares
df = degrees of freedom

#3: Find the t-statistic (cont.)

4

$$t = \frac{\text{sample mean difference} - \text{population mean difference}}{\text{estimated standard error}} = \frac{(\overset{93}{M_1} - \overset{85}{M_2}) - \overset{0}{(\mu_1 - \mu_2)}}{\boxed{s_{(M_1-M_2)}}\ 2}$$

REMEMBER!

M = sample mean
s² = sample variance
s²ₚ = pooled variance
n = sample size
SS = sum of squares
df = degrees of freedom

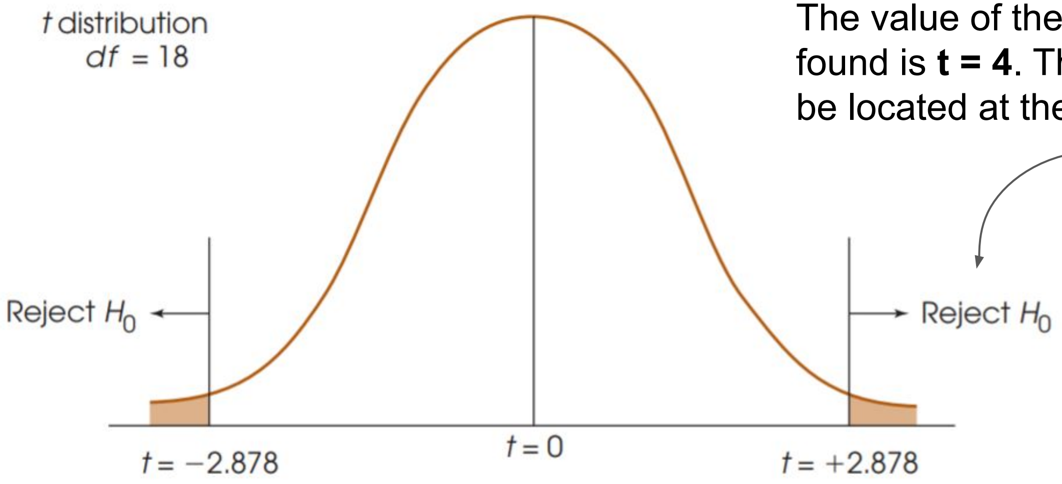
20

$$s_{(M_1-M_2)} = \sqrt{\frac{s_p^2}{\underset{10}{n_1}} + \frac{\boxed{s_p^2}}{\underset{10}{n_2}}}$$

$$\text{pooled variance} = s_p^2 = \frac{\overset{200}{SS_1} + \overset{160}{SS_2}}{\underset{9}{df_1} + \underset{9}{df_2}}$$

#4: Locate the t-statistic

The t distribution for **df = 18** is and **α = 0.01**, the critical region consists of the extreme **1% of the distribution** and has **boundaries** of **t = +2.878 and t = -2.878**.



The value of the **t-statistic** that we found is **t = 4**. This value of t should be located at the **critical region**.

We can be confidence to **reject the H_0**. Therefore, we can conclude that there **is a significant difference** between the two groups.

#5: Calculate the effect size

$$\text{estimated } d = \frac{\text{estimated mean difference}}{\text{estimated standard deviation}} = \frac{M_1 - M_2}{\sqrt{s_p^2}}$$

$$d = \frac{M_1 - M_2}{\sqrt{s_p^2}} = \frac{93-85}{\sqrt{20}} = \frac{8}{4.7} = \boxed{1.79}$$

SMALL Effect Size

How **huge** is the effect size?

#6: Reporting the result

APA Style:

The students who study using **experiment-based learning had higher** high school **grades** (M=93, SD=4.71) than the students who study using **reading-based learning** (M=85, SD=4.22). The **mean difference was significant**, t(18)=4.00, p <0.01, d=1.79.

Other way:

The difference was significant, t(18)=4.00, p=0.01, d=1.79.

What can we conclude?

Thank you!

Let's connect!

Email: bahyhelmi97@gmail.com
GitHub: https://github.com/bahyhelmihp
LinkedIn: https://linkedin.com/in/bahy