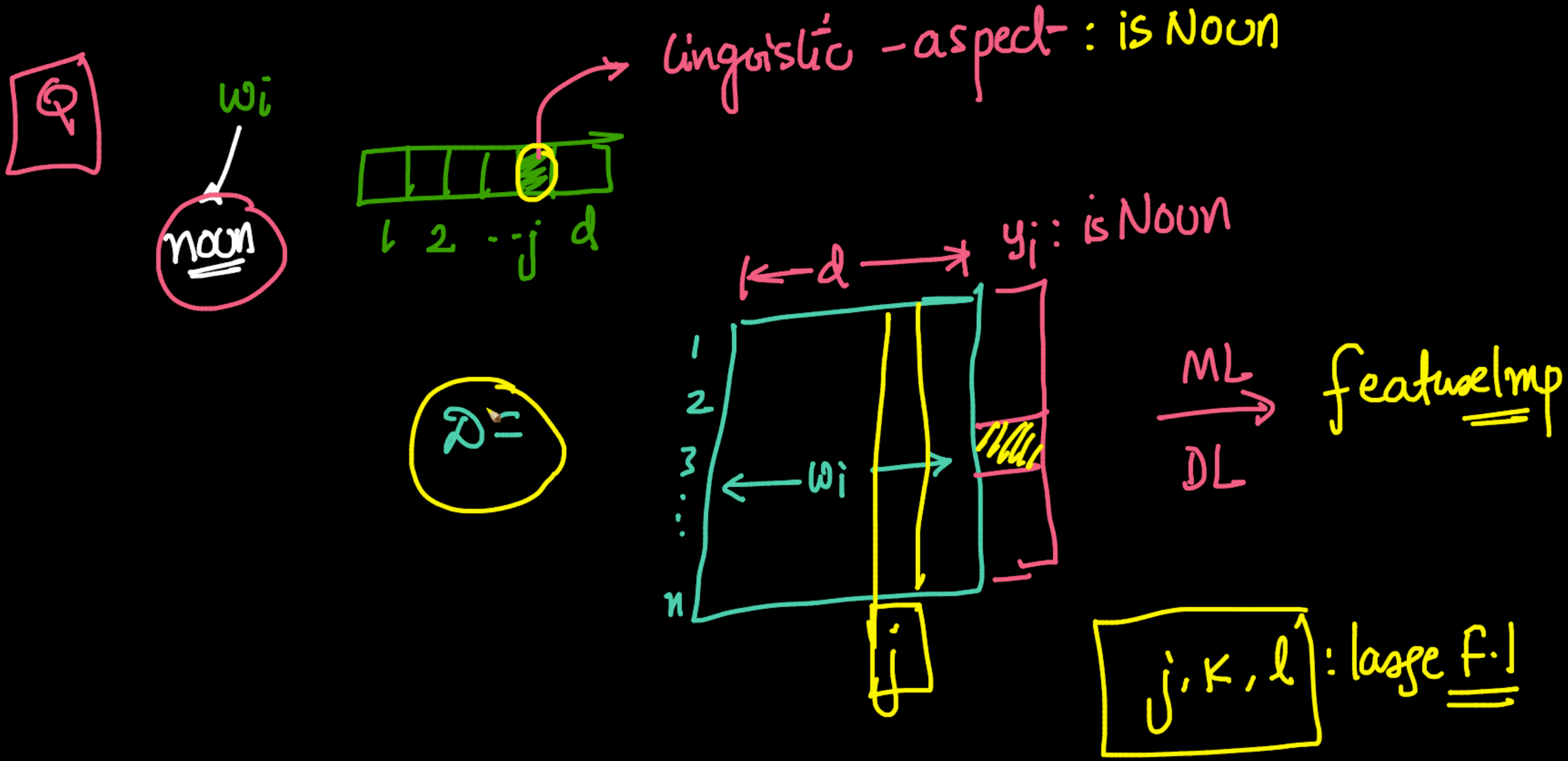

linguistic -aspect : is Noun
wi
noun
1 2 .. j d
yi : is Noun
d
1
2
3
n
wi
j
D=
ML
DL
featureImp
j, K, l : large f.I

Session - 2

Classical Hypothesis - tests

- z-test ; T-test : Z-proportions test
- χ^2-test

Next: → ANOVA & variations

Questions → questions; upvote (topic)

→ administrative ...

↳ email, phone

Framework of
Hyp-testing

✓ 1. Data; Task; Define H_0, H_a carefully

✓ 2. Expt design/use a 'sensible' Test-statistic (T)
conduct : $\underline{T_{obs}}$

3. Using T & H_a & H_0 : one-side or 2-side test

Chrome File Edit View History Bookmarks Profiles Tab Window Help

4 $P(T \text{ is as extreme as } T_{obs} \mid H_0) = \text{p-val}$

disb of T under H_0

5 $\alpha = 5\%$ (default) : agree

6 reject H_0 or accept H_0

Z-test
T-test
$\boxed{\text{KS-test}}$ ✓
χ^2-test
Z-prop test
ANOVA

5 / 5

Q

H_0: coin is fair

H_a: coin is biased

T: # tails in 500 tosses

✓ $\boxed{\text{2-sided}}$ or 1-sided (R/L)

Pfizer

medicines for covid : M_1 | M_2

100 | 90

randomly split

#days to recover : $x_{11}, x_{12} \dots x_{1,100}$ | $x_{21}, x_{22} \dots x_{2,90}$

mean #days to recover

m_1 M_2

[Compare means of 2-sample]

can

Permutation test : large sample
no assumptions on data

Z-test/T-test : small-samples
data meets assumptions

Preferred

More powerful

Finder File Edit View Go Window Help 10:28

Sample-means

Z-test

↳ std. normal variable

1 (a) D : x_{1i} ; x_{2j} → M_1 & M_2

$i : 1 \to 100$; $j \to 1 : 90$

(b) Task: Compare mean recovery. times of M_1 & M_2

(c) $H_0 : M_1 = M_2$

$H_a : M_1 \neq M_2$

ideal : $\mu_1 = \mu_2$; : $\mu_1 \neq \mu_2$ } pop. means

10 / 11

Finder File Edit View Go Window Help

2

$$T_{obs} = \frac{\overbrace{M_1 - M_2}^{\text{sample Means}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ → pop. std. dev

pop. finite

Mean & σ ✓

↓

sample-means

Z-test-assumptions

→ M_1 & M_2 are Normally disb : CLT

→ std-dev σ_1 & σ_2 are known → If n_1 are n_2 are reasonally large (≥ 30)

$\sigma_1 \approx S_1$; $\sigma_2 \approx S_2$

11 / 12

$$T = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2 \uparrow}{n_1} + \frac{\sigma_2^2 \uparrow}{n_2}}} \uparrow$$

Ho is true → 0 (Close to)

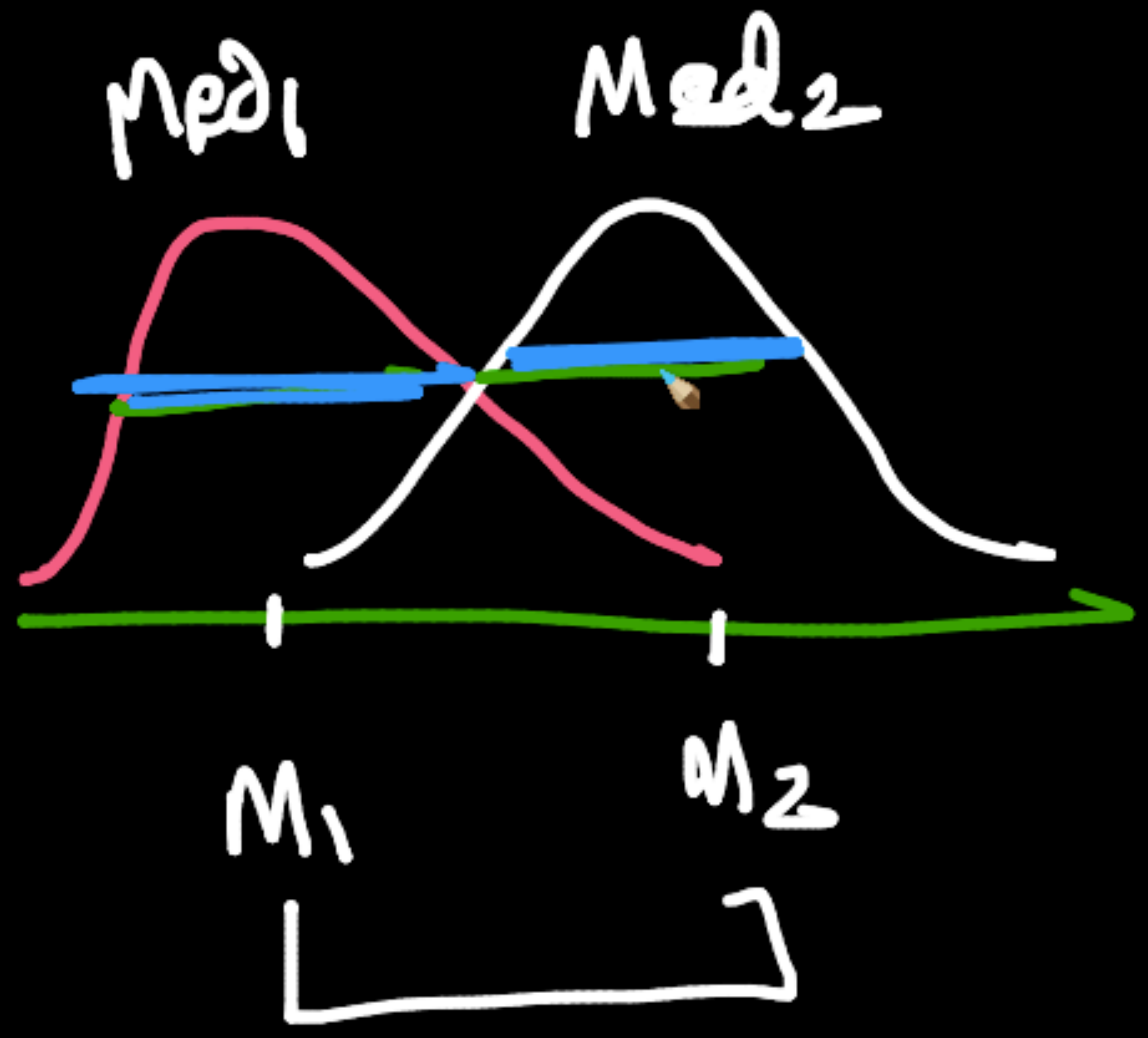
Ha is true → $M_1 > M_2$

→ $M_2 > M_1$

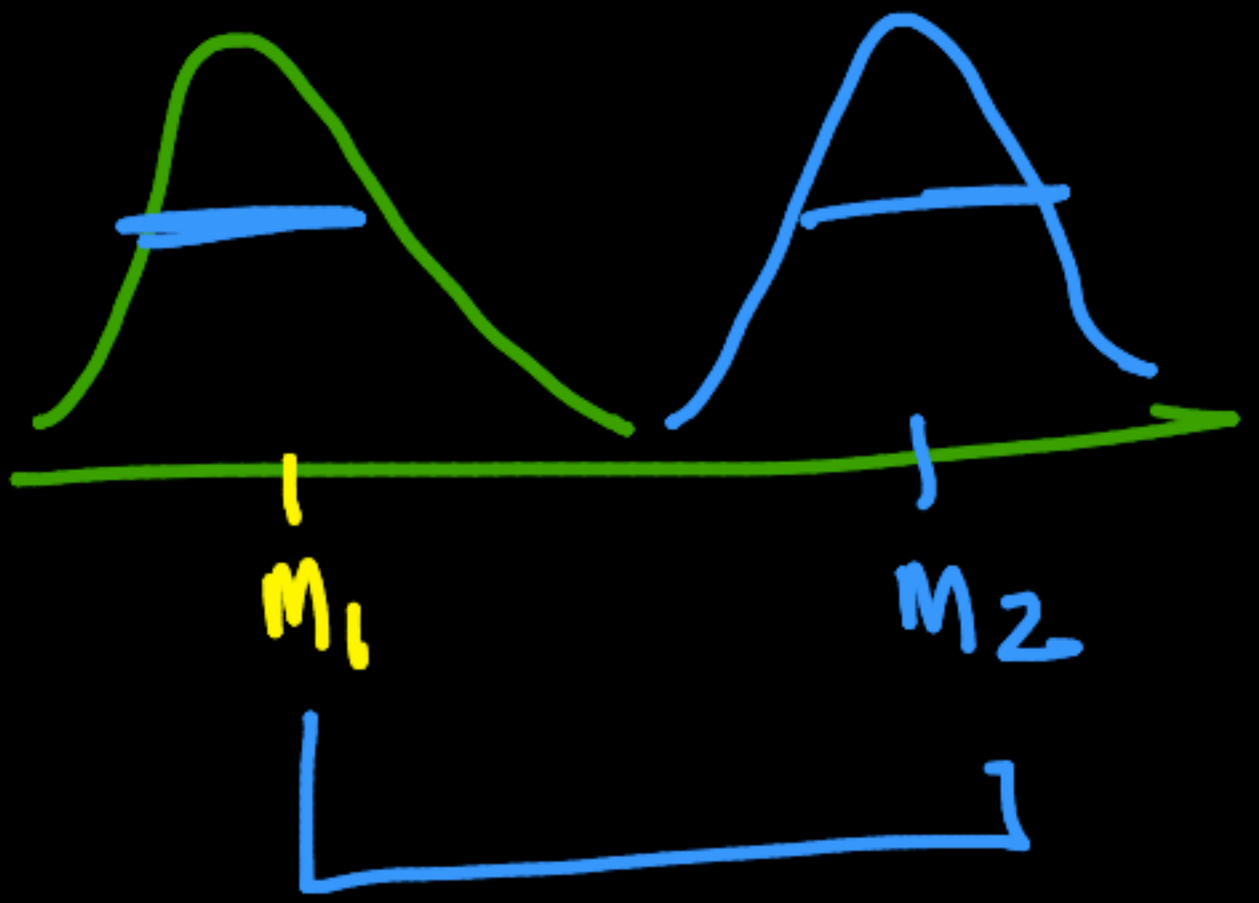
T will be farther away from 0 on either side

Let $M_1 - M_2 = 2$

σ_1 & σ_2 are large $\Rightarrow T \downarrow$

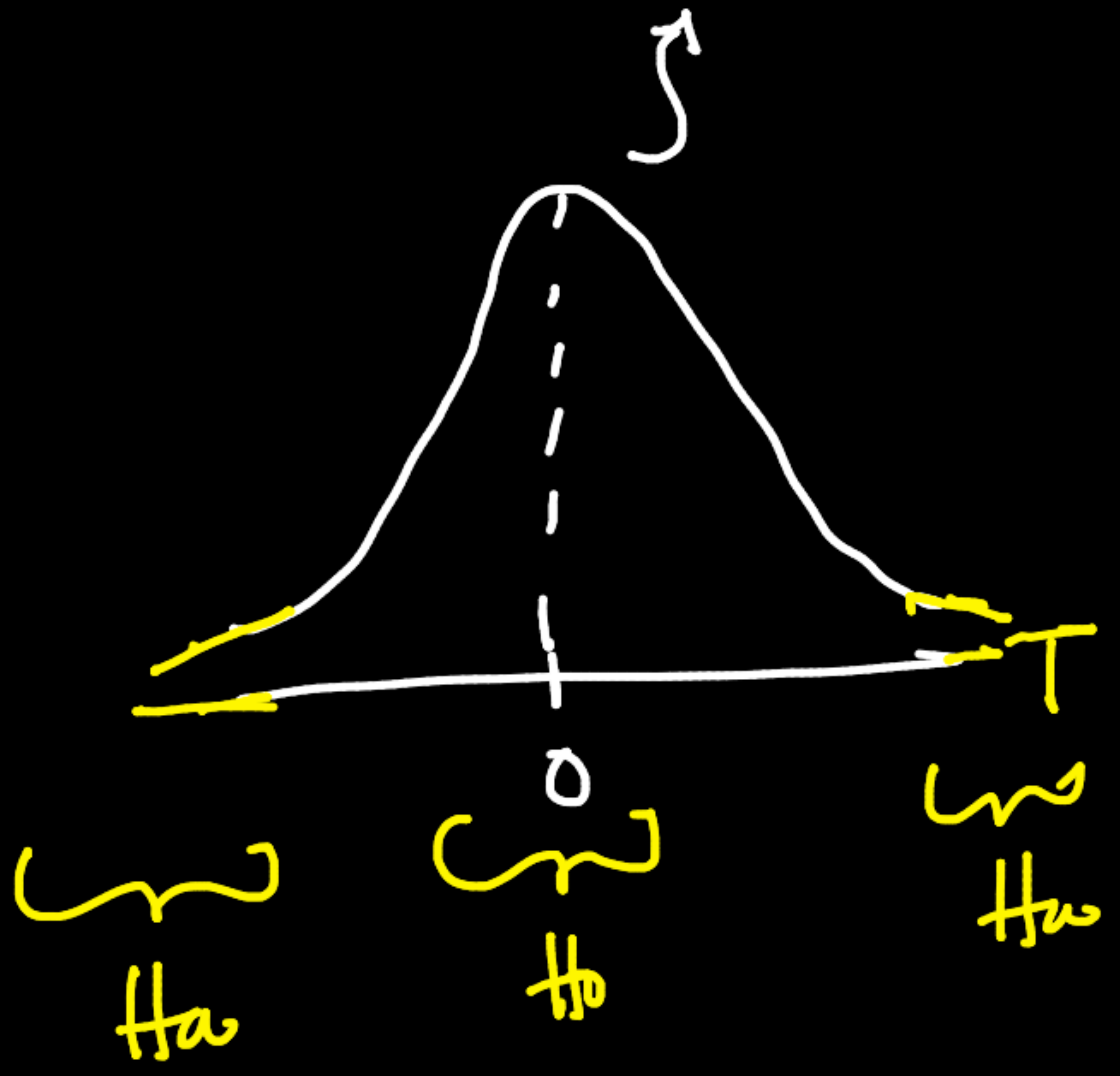


sample-means ~ Gaus



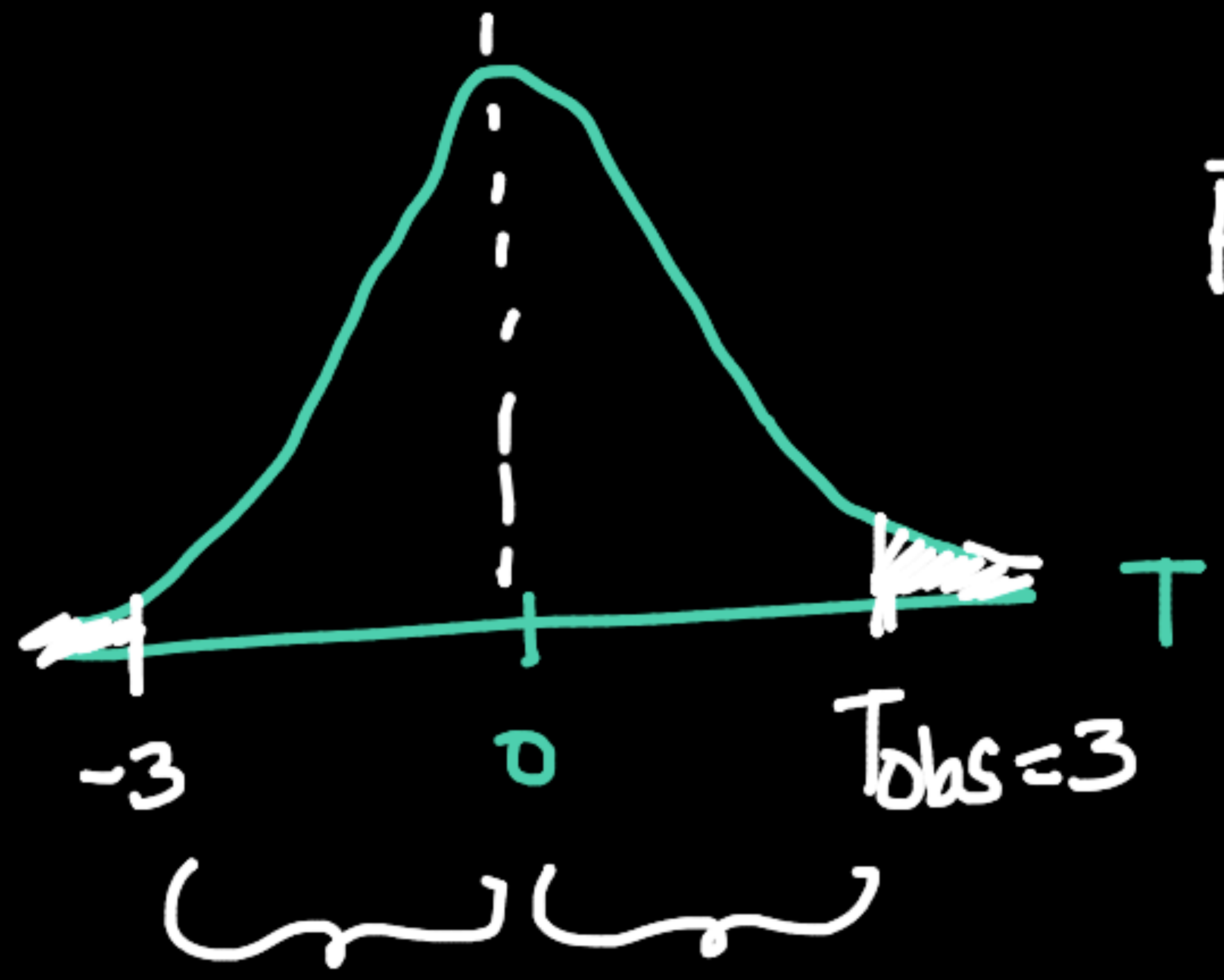
4

T under $H_0 \sim N(0,1)$



2-sided test

5 p-value:



p-val $=$ $= 0.3\%$

$= 0.3\%$

68-95-99.7% rule

6

$\alpha = 5\%$ (default)

$0.3\% = p\text{-val} < \alpha = 5\%$

$$\Downarrow$$

reject H_0 & accept H_a $\boxed{\mu_1 \neq \mu_2}$

Z test

→ Sample Means are normal under CLT

→ σ_1 & σ_2 are known or estimatable (n_1 & n_2 are large)

$$T = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$: 2-sided test

Q

Can we use the z-test to compare medians?

Not-always

Sample medians dist need not be gaussian (CLT)

Q

$n_1 < 30$; $n_2 = 50$

Z-test ?

do not know σ_1 & σ_2

$\sigma_1 \neq S_1$

$\rightarrow$ $\boxed{\text{T-test}}$

$\downarrow$

list: $\boxed{\text{barley}}$

$\boxed{Q}$ [Med_1 Med_2 - - - Med_k

← ANOVA

1

medicine 1

$n_1 < 30$

(20)

medicine 2

$n_2 < 30$

($n_2 = 25$)

T-test

$\downarrow$

$\boxed{\text{science}}$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

[Compare means of 2-samples]

T-test

Assumptions

- ✓ m_1 & m_2 are normally dist (CLT)
- σ_1 & σ_2 are not known / not reasonably estimatable
- samples are random & indep of one another

  (↳ True for all tests)

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:58

$$\left\{ T = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right.$$

sample estimates

$(n_1 + n_2 - 2) \uparrow$

T-dist ~ Z-dist

T-test ~ Z-test

proof

$T \sim$ t-dist (ν)

$\nu = n_1 + n_2 - 2$

degrees of freedom

INTUITION: when data is small

Chrome File Edit View History Bookmarks Profiles Tab Window Help

Student t pdf - Student's t-dis

en.wikipedia.org/wiki/Student%27s_t-distribution#/media/File:Student_t_pdf.svg

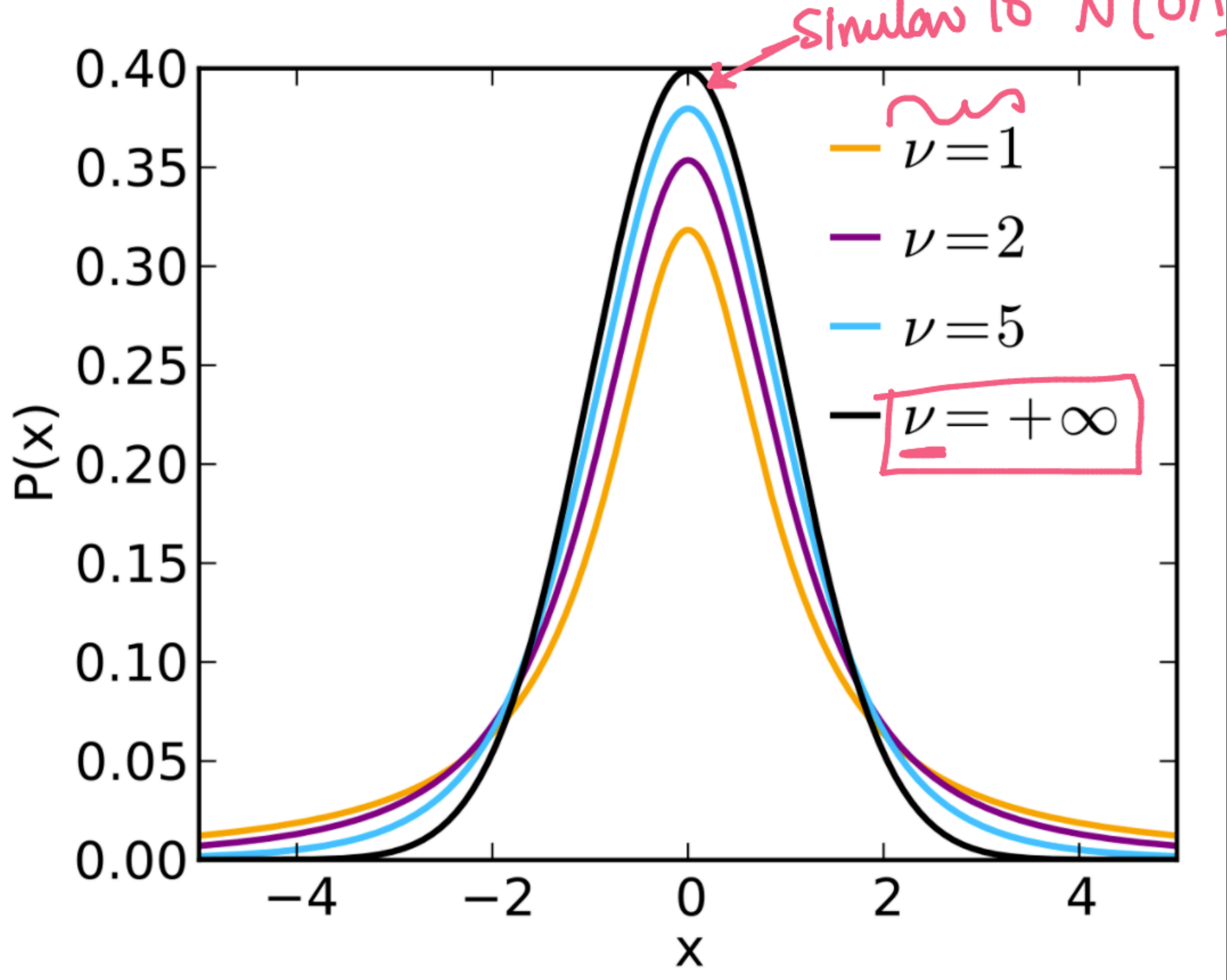
$\nu = n_1 + n_2 - 2$

↓

t-disb

$\approx$ z-disb

similar to N(0,1)



t-disb

Plot of the density function for several members of the Student t family.

More details

Skbkekas - Own work

25 / 25

Sample sizes are small $\xrightarrow{Y}$ T-test

$\xrightarrow{N}$ $\boxed{\text{z-test}}$ ✓

$\boxed{\text{Compare means}}$

Sample-means ~ Normal

Amazon:

(data)↑↑

means! → z-test → means

z-test → covid - drugs

means! → permutation test → medians

P99 . . .

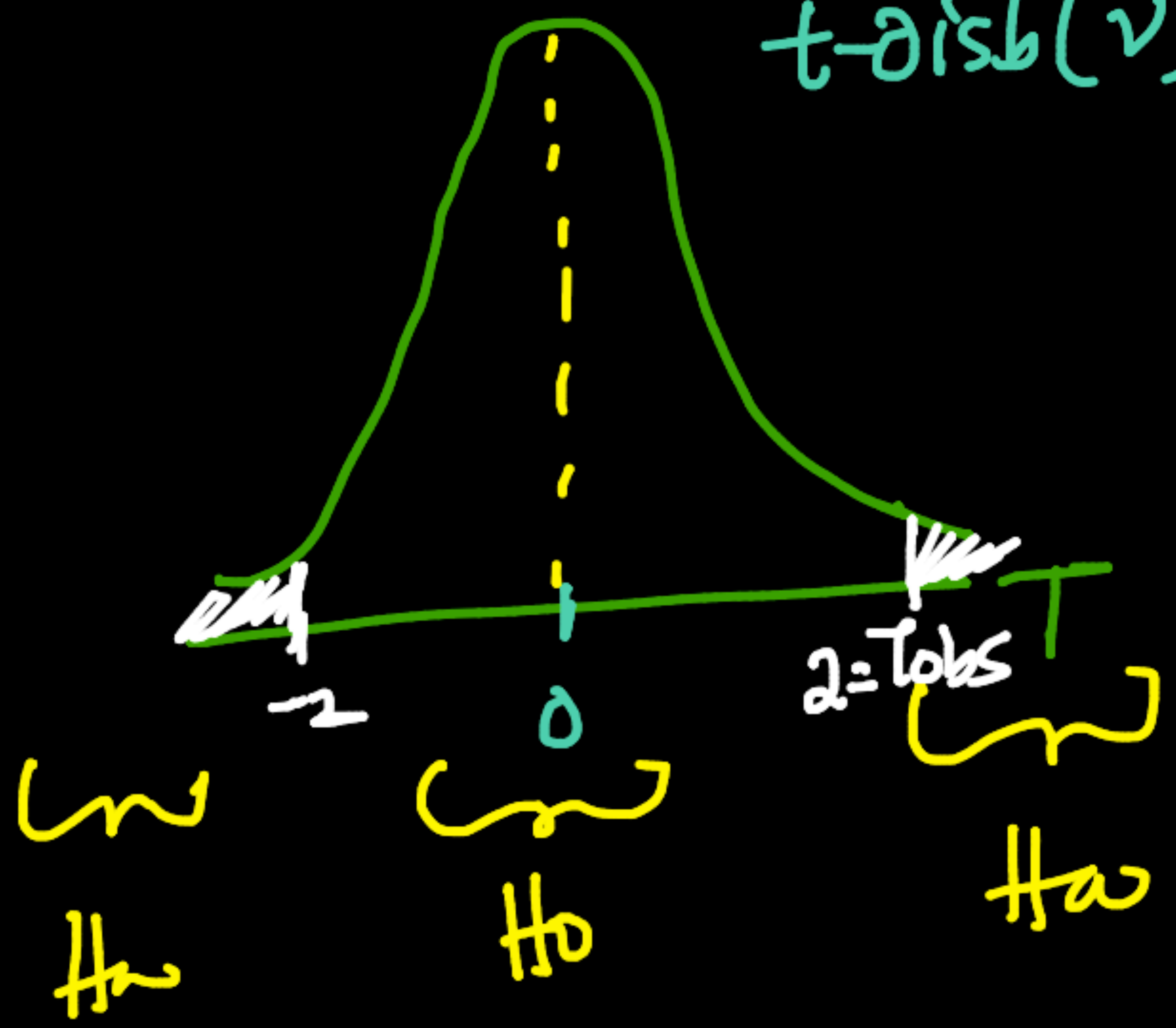
permutation test → non-param

χ^2-test

T-test (contd)

$$T = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t-dist(ν)



2-sided

☐ : p-val vs α

↓

accept or Reject Ho

Amazon:

Survey

Prime-sub

199 — 299

Z-test
or T-test

PM

Sample → Survey

Rs 5,000

Cost

100

199 : renew next Month

299 : renew ?

100

- 1-day delivery
- 1000+ movies . . .

11:11

- Z-prop-test
- Clarify

next

- χ^2 test
- ANOVA
- paired test; type1,2

z-proportions test

e-commerce

1 A: old page → $\dfrac{conv_1}{visits_1}$ = conv. rate = $\hat{p}_1$

10

40

2 B: new page → $\dfrac{conv_2}{visits_2}$ = $\hat{p}_2$

Sample ratios

Z-test & T-test : Comparing Means

Z-prop test : proportions comparision

$$H_0: \overbrace{p_1}^{\text{pop. ratios}} = p_2$$

$$H_a: p_1 \neq p_2$$

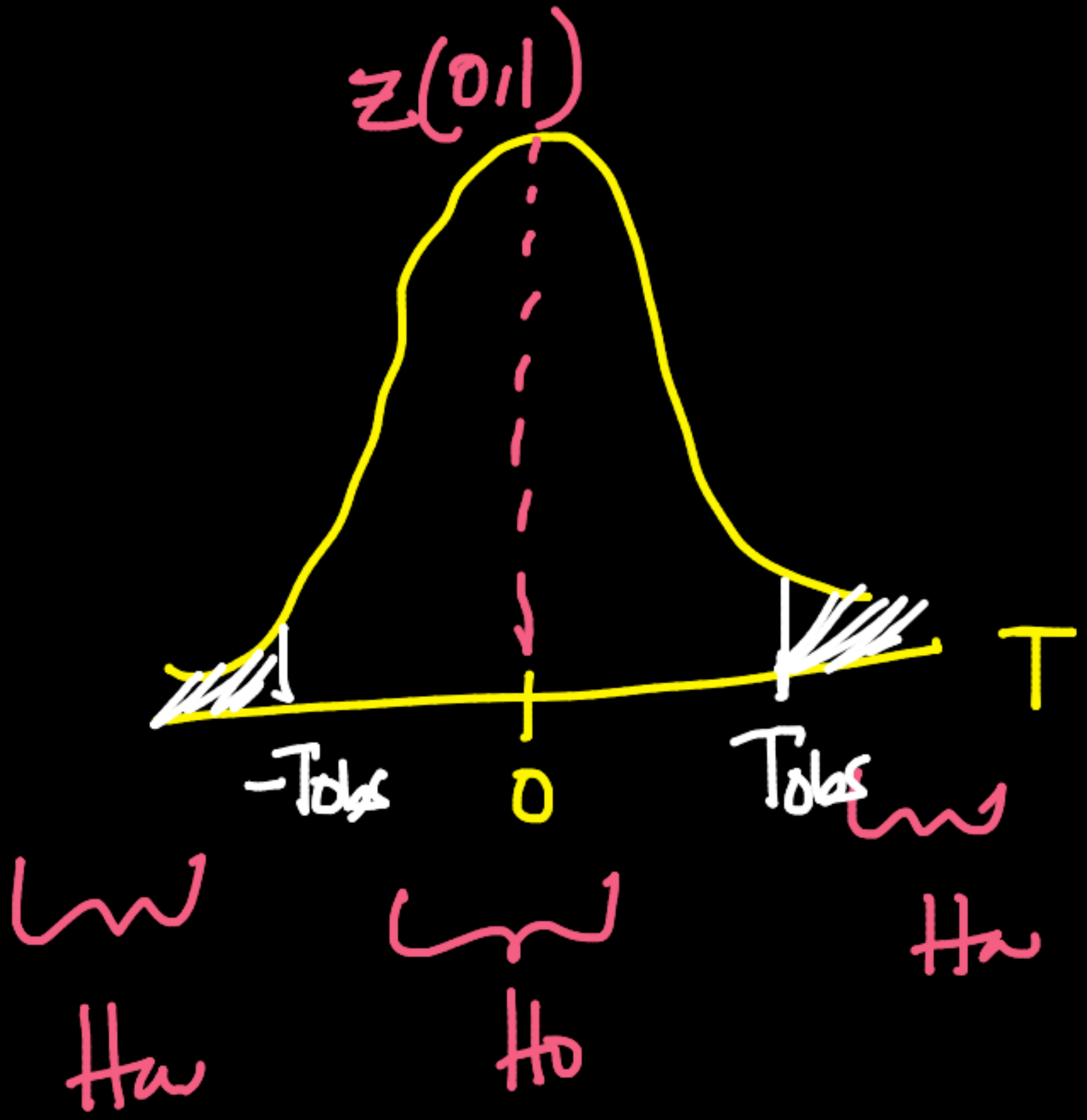
Samples > 30

n_1 & $n_2 > 30$

$$\boxed{T} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim Z(0,1)$$

$$\hat{p} = \frac{c_1 + c_2}{v_1 + v_2} \text{ : overall ratio}$$

T under H_0:



2-sided

▨ : p-val

p-val vs α

Med 1 → A / B → Setup

Med2

A / B → hyp-testing

hyp-testing → A / B

reject H_0 & H_a

(Math)

Permutation Testing:

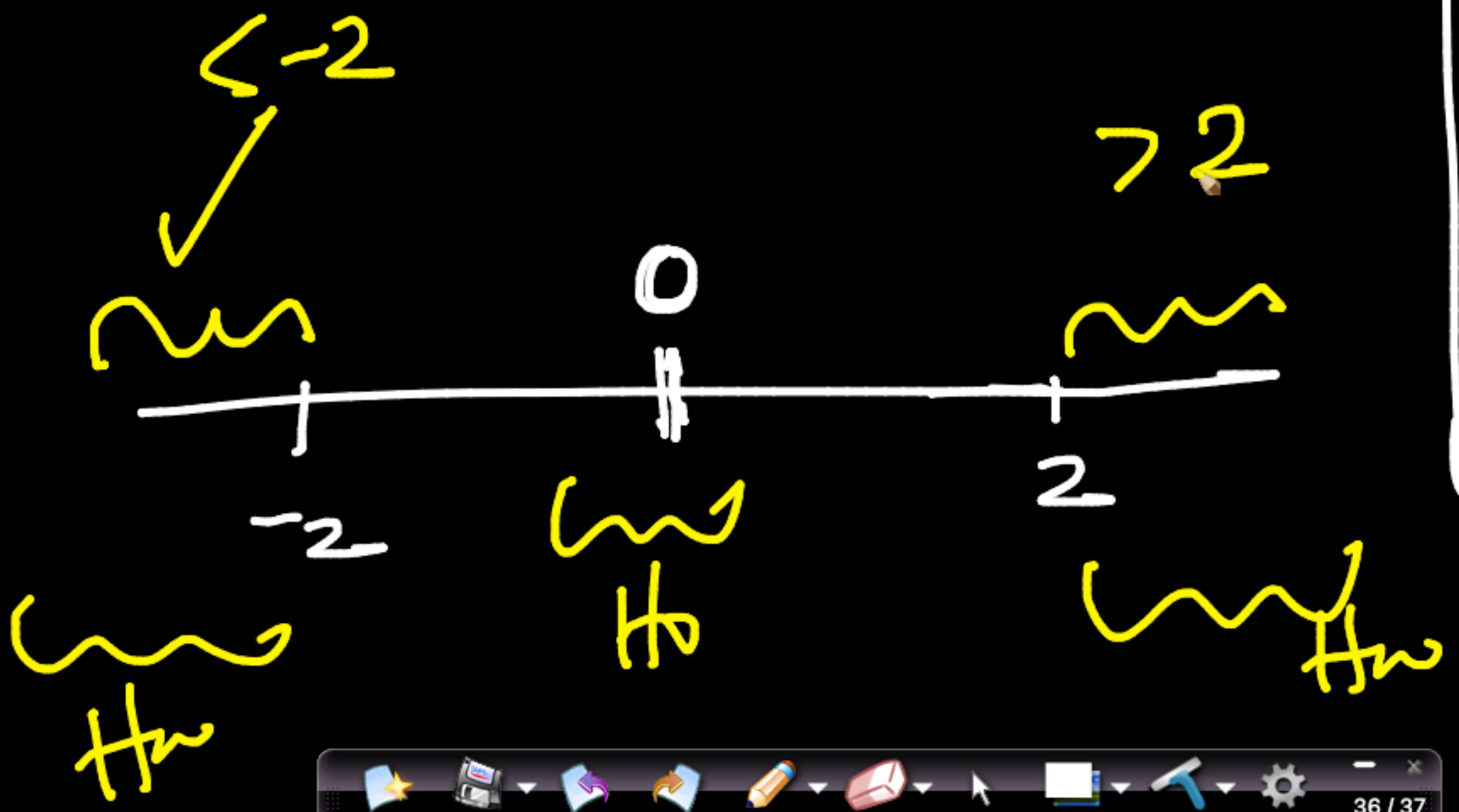
$\rightarrow \{ H_0: M_1 = M_2 ; \quad H_a: M_1 \neq M_2$

$T = \widetilde{M_1 - M_2}$

$T_{obs} = 2$

samples

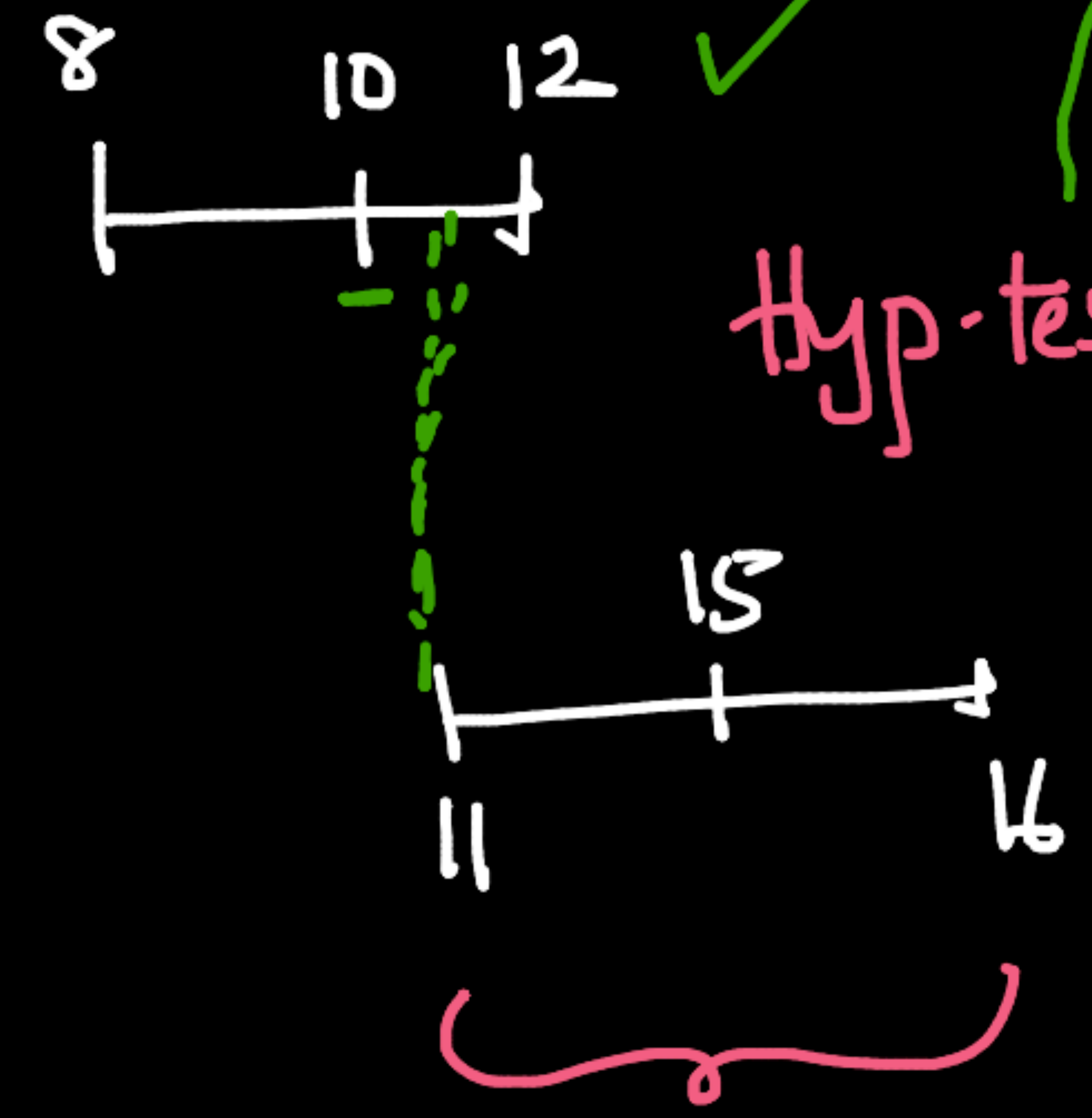
$\hookrightarrow \left\{ \begin{array}{c} T_1 \\ T_2 \\ \vdots \\ T_{10,000} \end{array} \right\}$ >2 or <-2



Medicine - recovery times

1 M_1

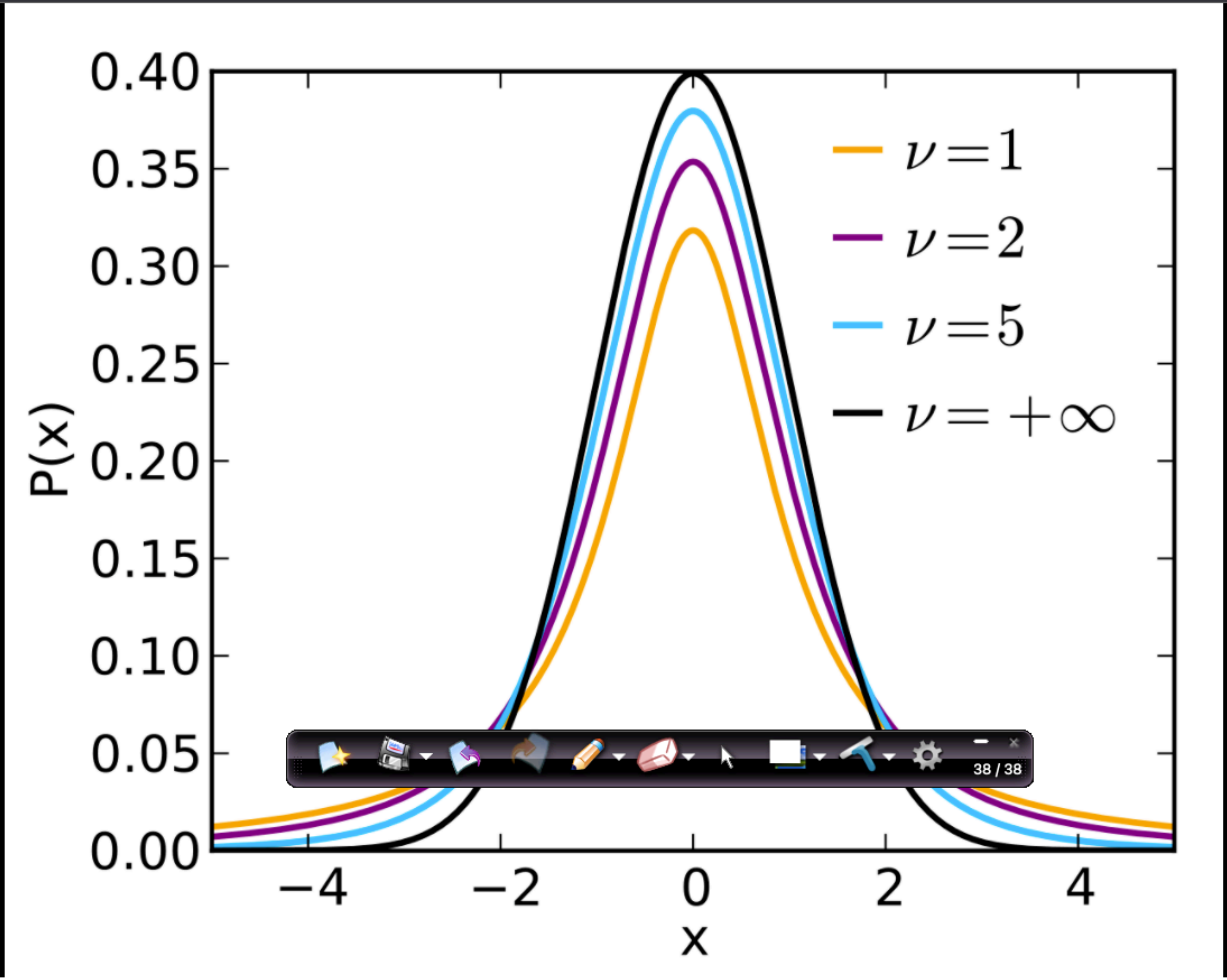
2: M_2



Hyp-test $\begin{cases} M_1 = M_2 \\ M_1 \neq M_2 \end{cases}$

95% C.I on rec times

[C.I → WHO; ICMR; pfizer ...



Plot of the density function for several members of the Student t family.

More details

Skbkekas - Own work