Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:03

Topics:

✓ ① Testing K-drugs recovery times → z/t-test

✓ ② [ANOVA → Analysis of variance — Intuition with framework]

③ ③ Test-statistic & F-dist^b

{ ④ Variations of ANOVA } ✓

[⑤ p-hacking

1 / 3

OPS

Questions → Q&A tab

→ "END: - - -

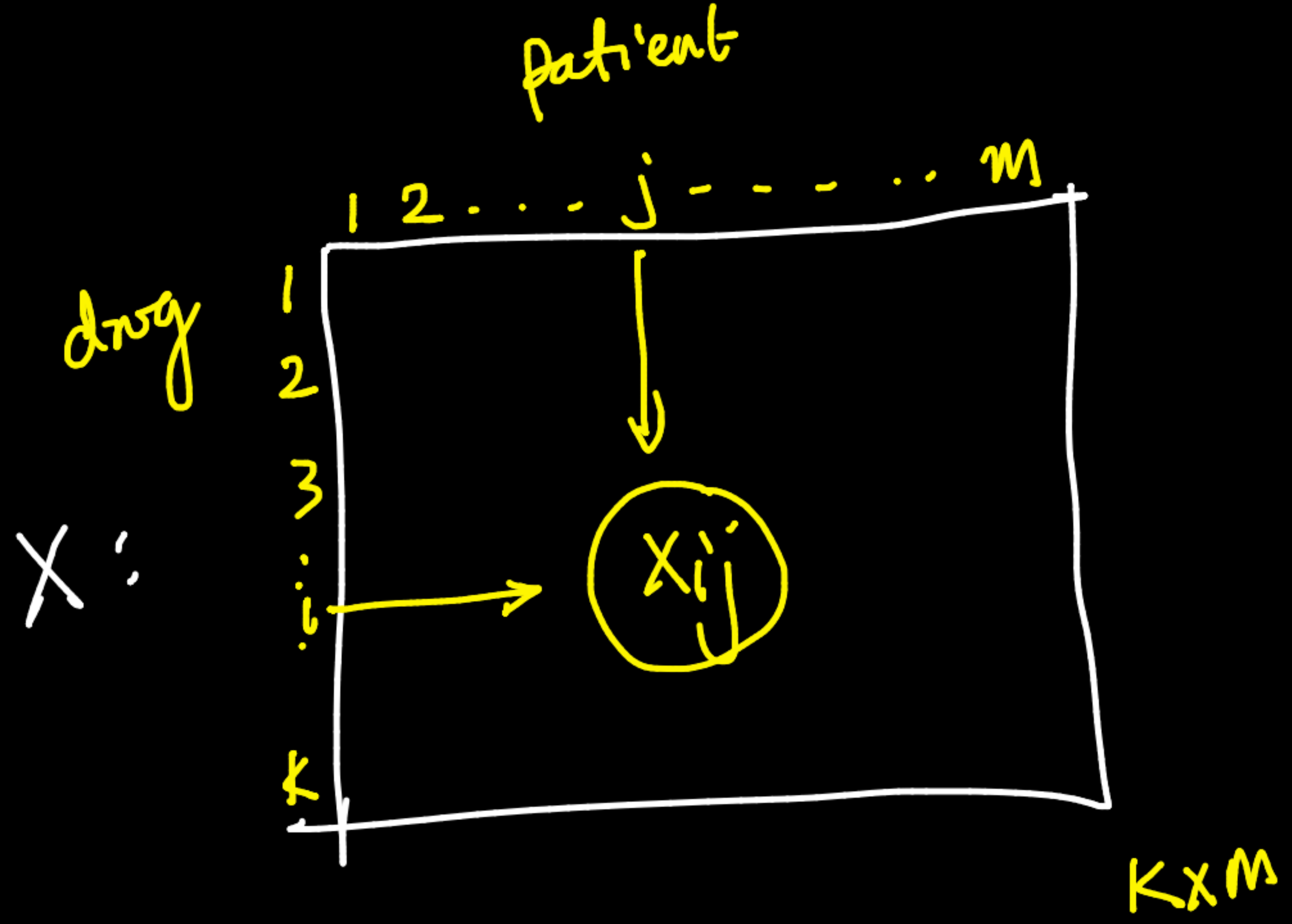
→ off-topic : "END ......

Chat → Y/N & interactivity

Problem:

Observed data:

- $\underline{K}$ (10) - drugs for covid → one-group [K-groups]
- m (100) - patients in each gp
- $n = m \times K$ (1000)
- X_{ij} = rec. time of j^{th} patient taking drug i


patient
1 2 . . . j - - - . . m
drug
1
2
3
i
k
X :
X_{ij}
K x M

$\boxed{Q}$ do all these k-drugs have similar mean (pop) rec. times or not

$$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k \quad ?$$

or (not)

✓ each pair of drugs

↳ if $m \geq 30$ → z-test

→ if $m < 30$ → T-test

1 2 3 5 K

C → K

T → (K-1)

3 vs 5
5 vs 3

K-drugs

How many test do we have to do

$$\frac{K \times (K-1)}{2} = {}^{K}C_2$$

Multiple HT

$K = 10$

$^{K}C_2 = \frac{10 \times 9}{2} = \boxed{45}$ hyp-tests

Any issue with this

$\downarrow$

X comp. expensive

$\alpha = 5\%$ (let)

↑ accept Ha incorrectly (EPR)

↑ rejecting Ho incorrectly

each test $\longrightarrow$ $\boxed{5\%}$ prob of FP (error)

$\boxed{Q}$

45 tests

Prob that we incorrectly reject H_0 (FP-error) in atleast one of the 45 tests

$$E \sim \text{Bin}(n = \underline{45},\ p = 0.05)$$

$$P(E \geq 1) = 1 - P(E = 0)$$

$$= 1 - {}^{45}C_0 \; 0.95^{45} \; 0.5^{0} \approx \boxed{90\%}$$

Commit FP error

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:25

Lesson: Multiple Hyp.test even @ small α can result in a high prob of FP-error

11 / 12

Q

HT1 → $\boxed{\mu_1 = \mu_2}$ → p-val, α

HT2 → $\boxed{\mu_2 = \mu_3}$ → α, p-val

$\Rightarrow \boxed{\mu_1 = \mu_2 = \mu_3}$

μ_1 vs μ_3

$\boxed{Q}$

Multiple hyp. Tests → FP-error ✓

↳ power $1-\beta$

$\boxed{\beta: \text{TN-error}}$

Q

observed

$$\text{HP-Test } 1: \quad \overbrace{\mu_1 = \mu_2} \rightarrow \text{p-value}$$

$$P\left(\begin{matrix} T > T_{obs} \\ T < -T_{obs} \end{matrix} \,\middle|\, H_0\right)$$

$$2: \quad \mu_1 = \mu_3$$

$\vdots$

14 / 15

Chrome File Edit View History Bookmarks Profiles Tab Window Help

Q

10 groups

45 Hyp. tests

Strict restriction on each HT ✓

$\alpha = 0.0001$

V.V small

$1 - (1-\alpha)^{45} = 5\%$

FP - error rate

$-(1-\alpha)^{45} = -0.095$

$45 \log(1-\alpha) = \log 0.095$

15 / 16

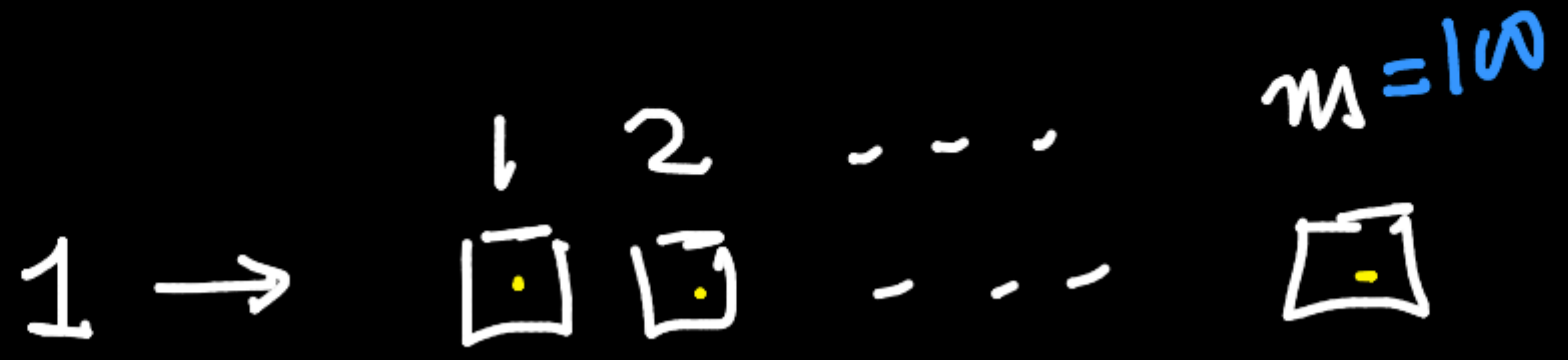
ANOVA

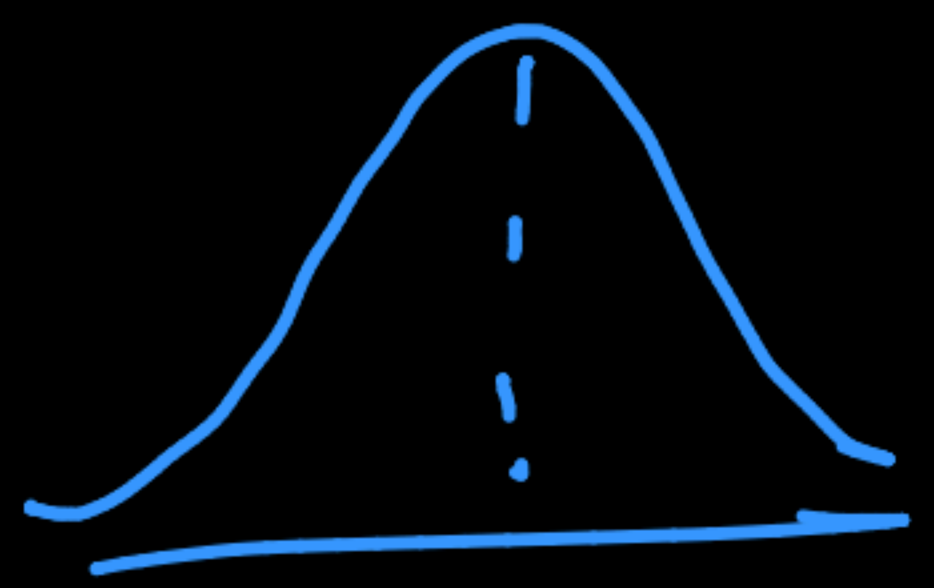
0.1% ✓

1%

[Single Test with $\alpha = 5\%$ (default) to compare means of K groups]

{assumption: each group follows gaussian Dist [strong assumption]





if not gaussian

↓

transform (box-cox)

↓

ANOVA

Practice:

non-gaussian $\xrightarrow{\text{Transform}}$ Close to gaussian

↓

ANOVA

Very different from Gaussian

↓

ANOVA (~~non-param...~~)

ANOVA

1. H_0: $\overbrace{\mu_1 = \mu_2 = \cdots = \mu_K}$

no difference in group means

H_a: there is some difference in atleast one of the gp's means

$\mu_1 \neq \mu_2 = \mu_3 = \cdots = \mu_K$

$\mu_1 \neq \mu_2 \neq \mu_3 = \mu_4 = \mu_5 \cdots = \mu_K$

Chrome File Edit View History Bookmarks Profiles Tab Window Help

② Test-statistic = $f = \cdots\cdots$ eqn.. (later)

↳ variance

Under H_0: $f \sim$ f-distb$(k-1, n-k)$

$k = \#$gps

$n = m \times k$

↳ Total #obs

ANOVA

↳ F-test

21 / 21

$$\text{Test-stat} = f = \frac{\text{mean-ss-distance between groups}}{\text{mean-ss-distances within groups}} = \frac{MSB}{MSW}$$

overall mean

$$\left\{ \sum_{i=1}^{K} \left(\bar{X}_i - \bar{X} \right)^2 \right\} \Big/ (K-1)$$

each gp's sample mean

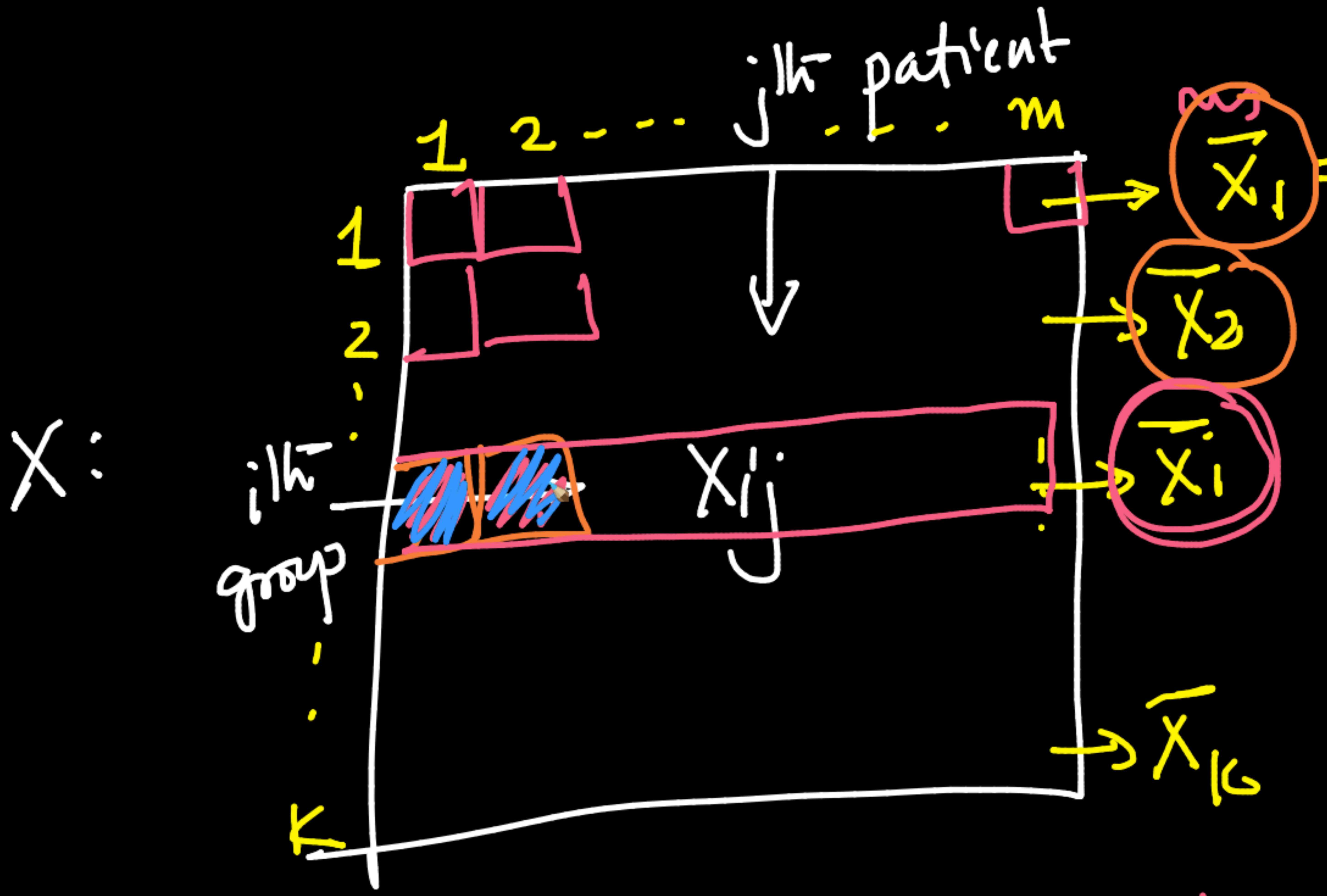
variance amongst $\bar{X}_i$'s

Test-stat

$$f = \frac{1}{n-K} \sum_{i=1}^{K} \sum_{j=1}^{m} \left(X_{ij} - \bar{X}_i \right)^2$$

avg across the K-gps

variance within each gp



$\bar{X}_1$ = sample Mean of rec. times of gp-1

$n = m \times K$

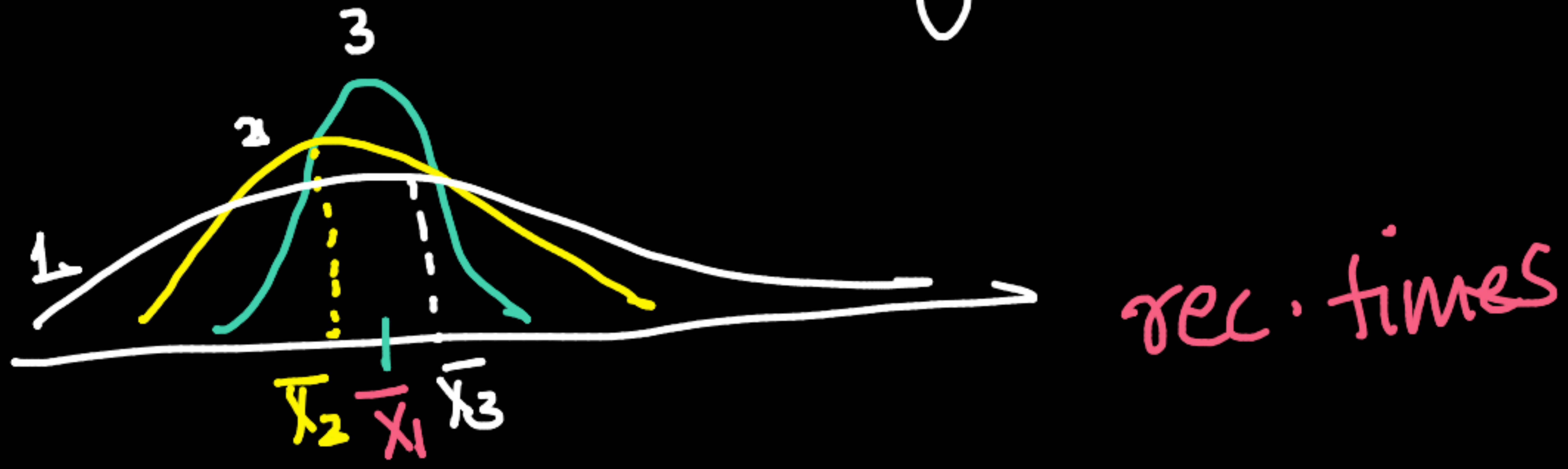
overall mean $= \bar{X} =$ sample mean of all n obs

Under H_0 : numerator $\rightarrow$ $\left\{ \begin{array}{l} \text{+ve always} \\ \text{zero} \end{array} \right.$ ~~very small~~

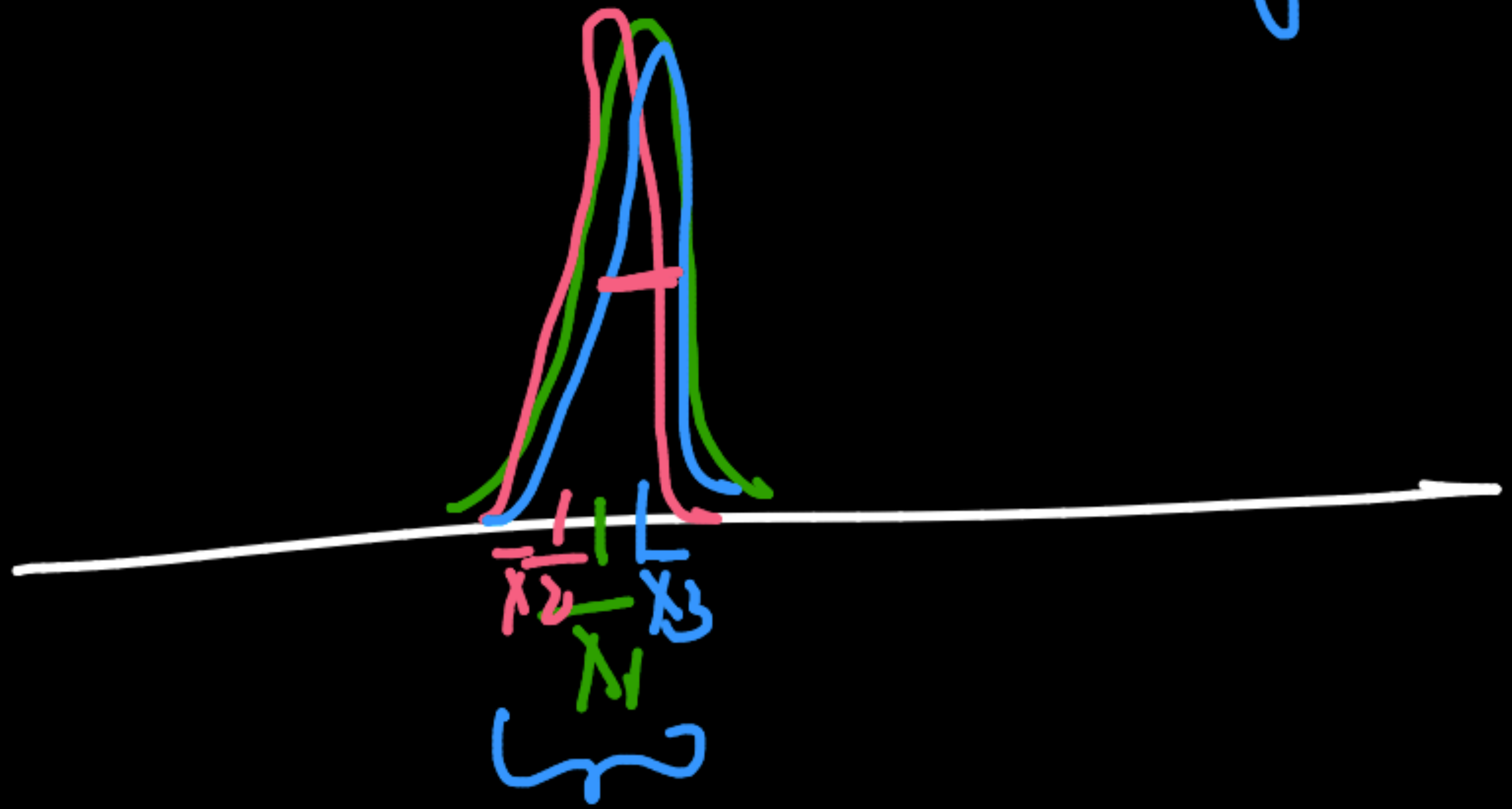
Under H_0: denominator → +ve high

average variance in each gp

Case 1:

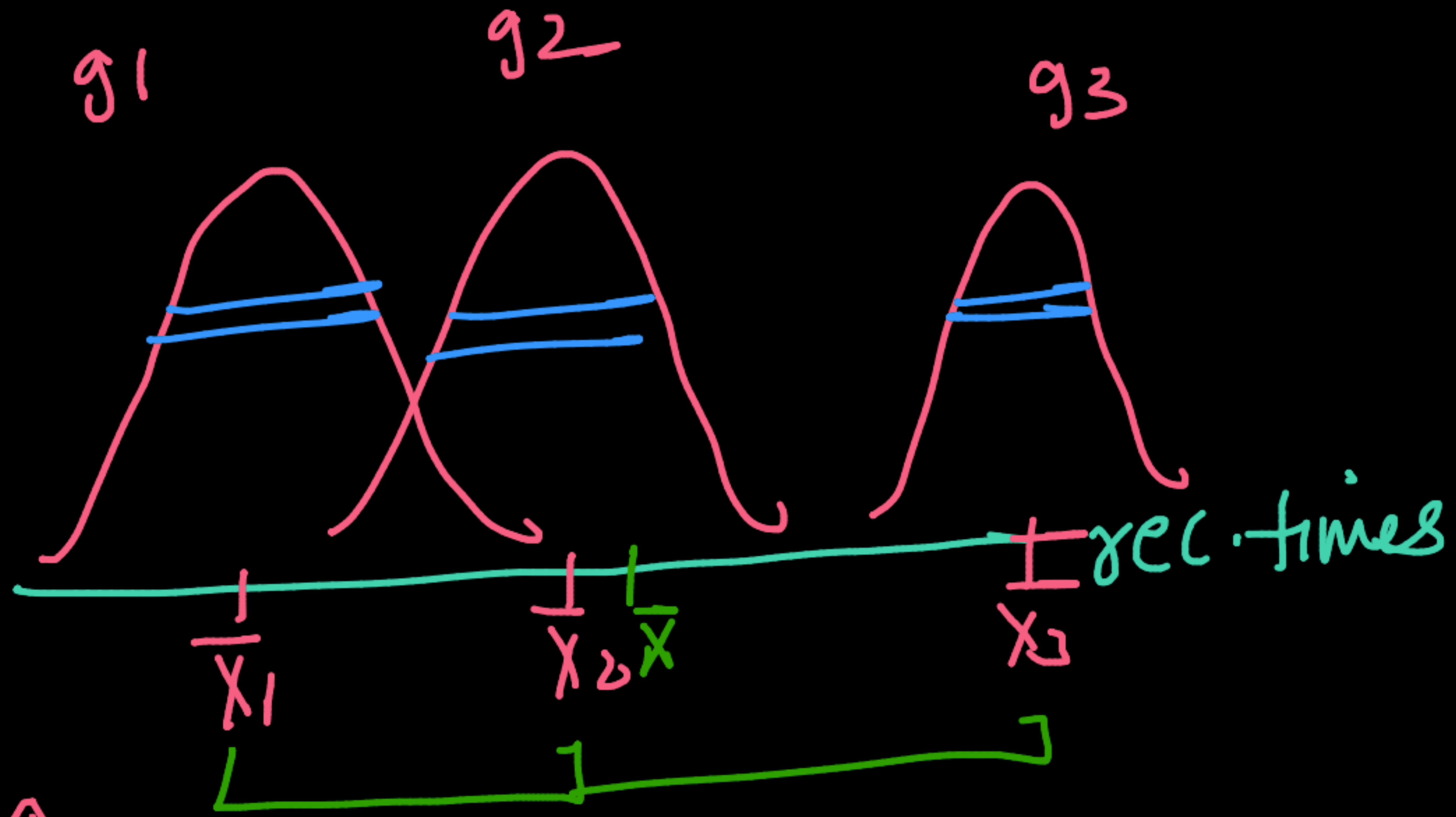


Case 2:-



$$f = \frac{MSB \downarrow}{MSW \downarrow}$$

Case1:

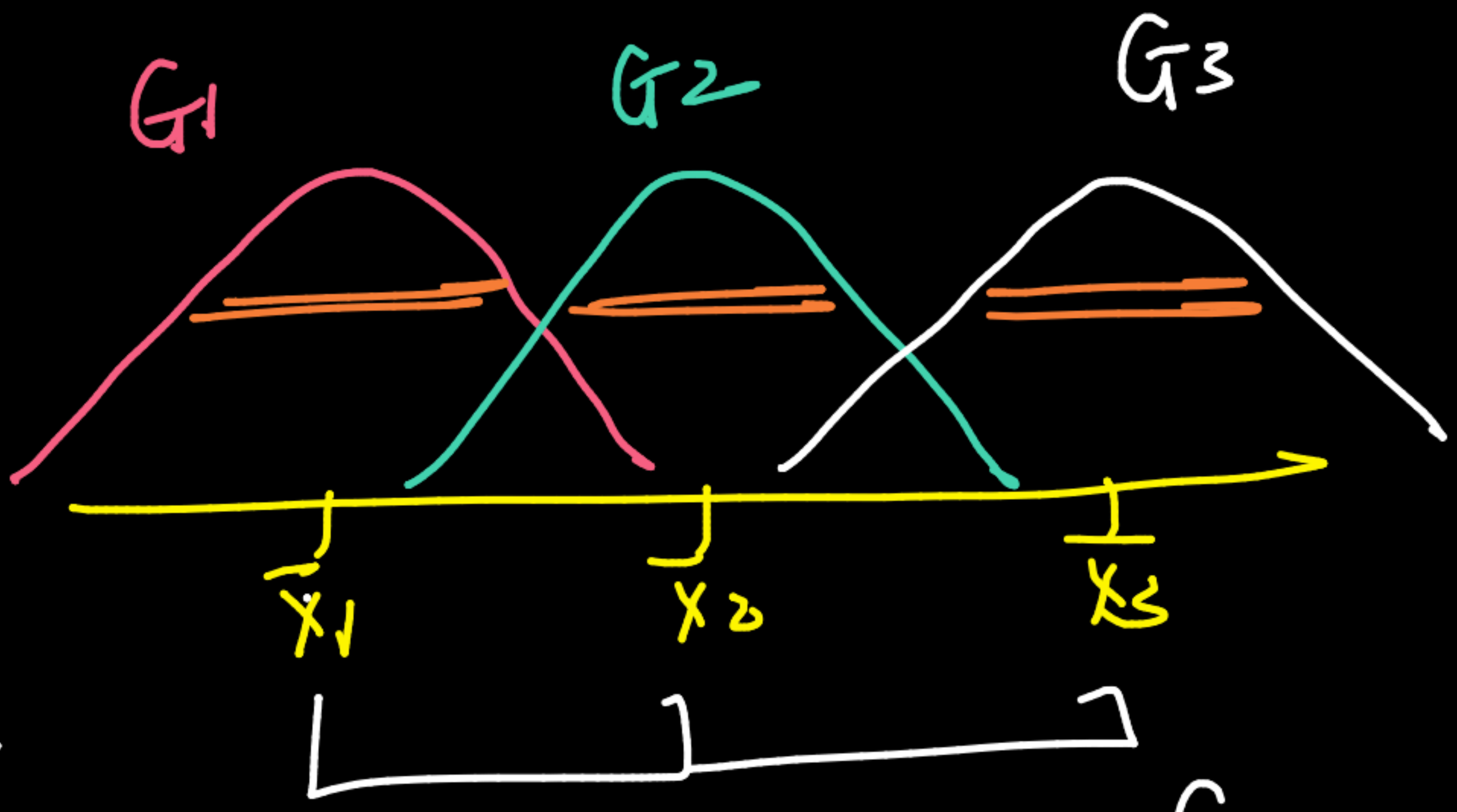


$$F_1 = \frac{MSB_1}{MSW_1} \uparrow$$

Case 2:

$MSB_2 = MSB_1$

$MSW_1 < MSW_2$



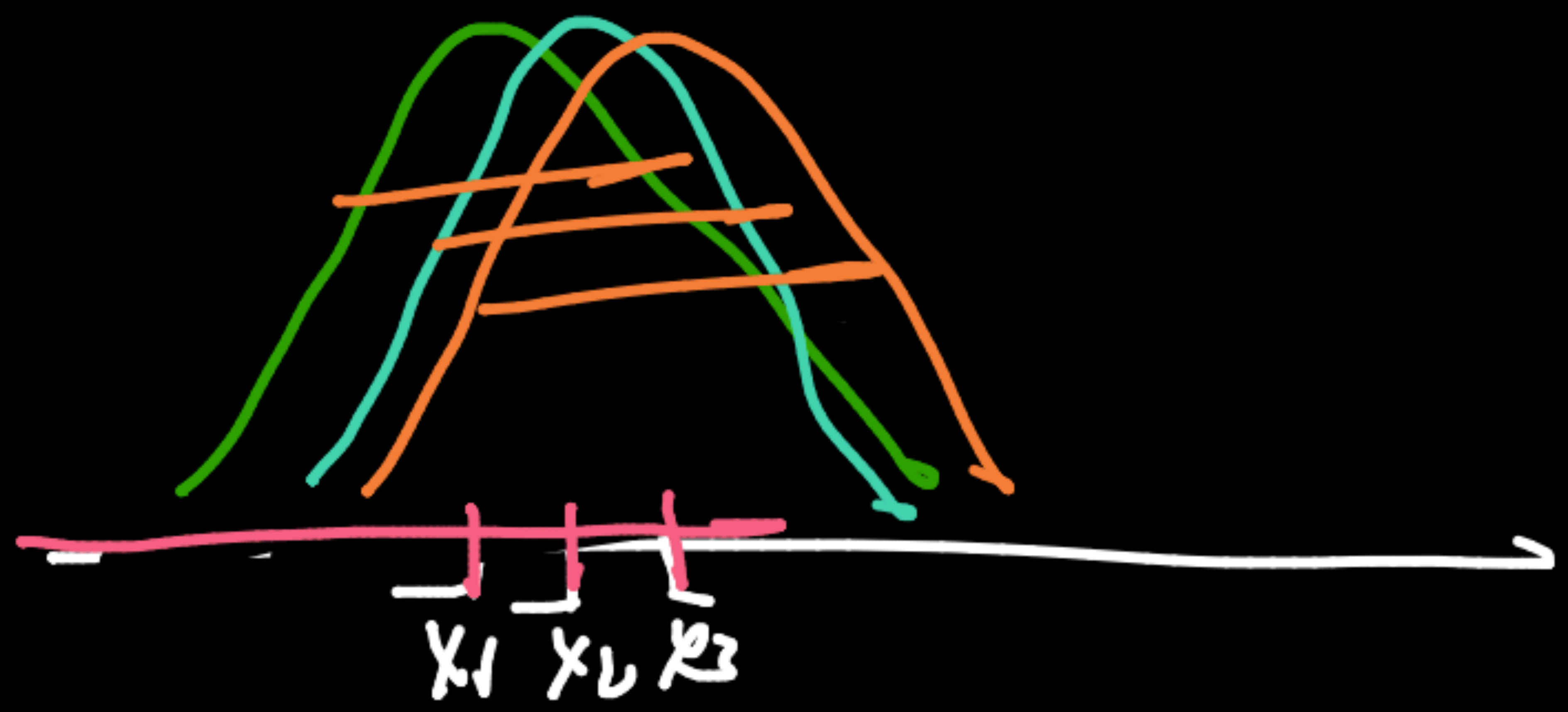
$$f_2 = \frac{MSB_2 \uparrow}{MSW_2} =$$

f_1 vs f_2

$\boxed{f_1 > f_2}$

Case 3:

each gp's var



$f_3 = \frac{MSB_3}{MSW_3}$ ↓ $\approx MSW_1$

f_1 vs f_3

$f_1 > f_3$

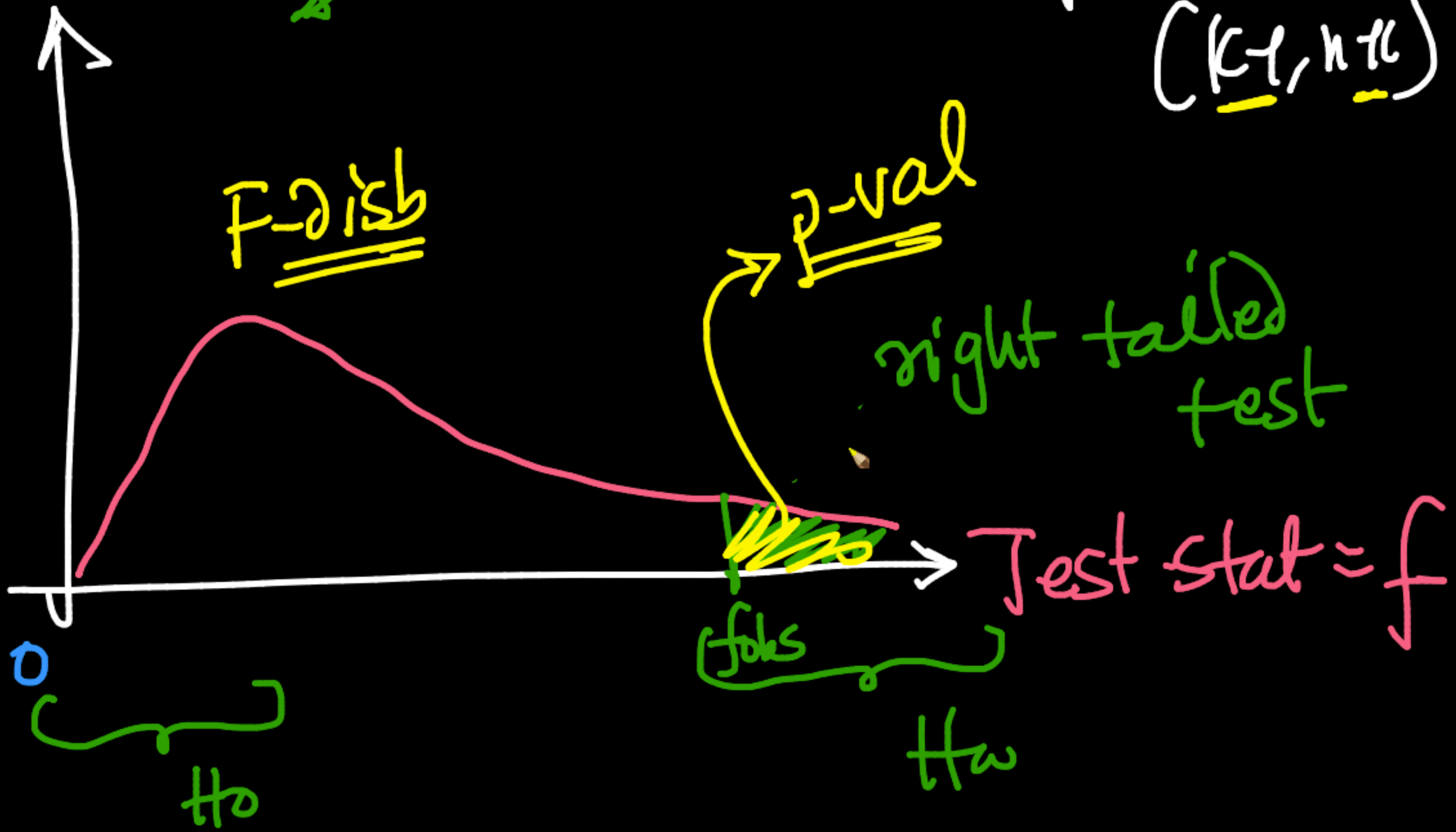
Test-stat

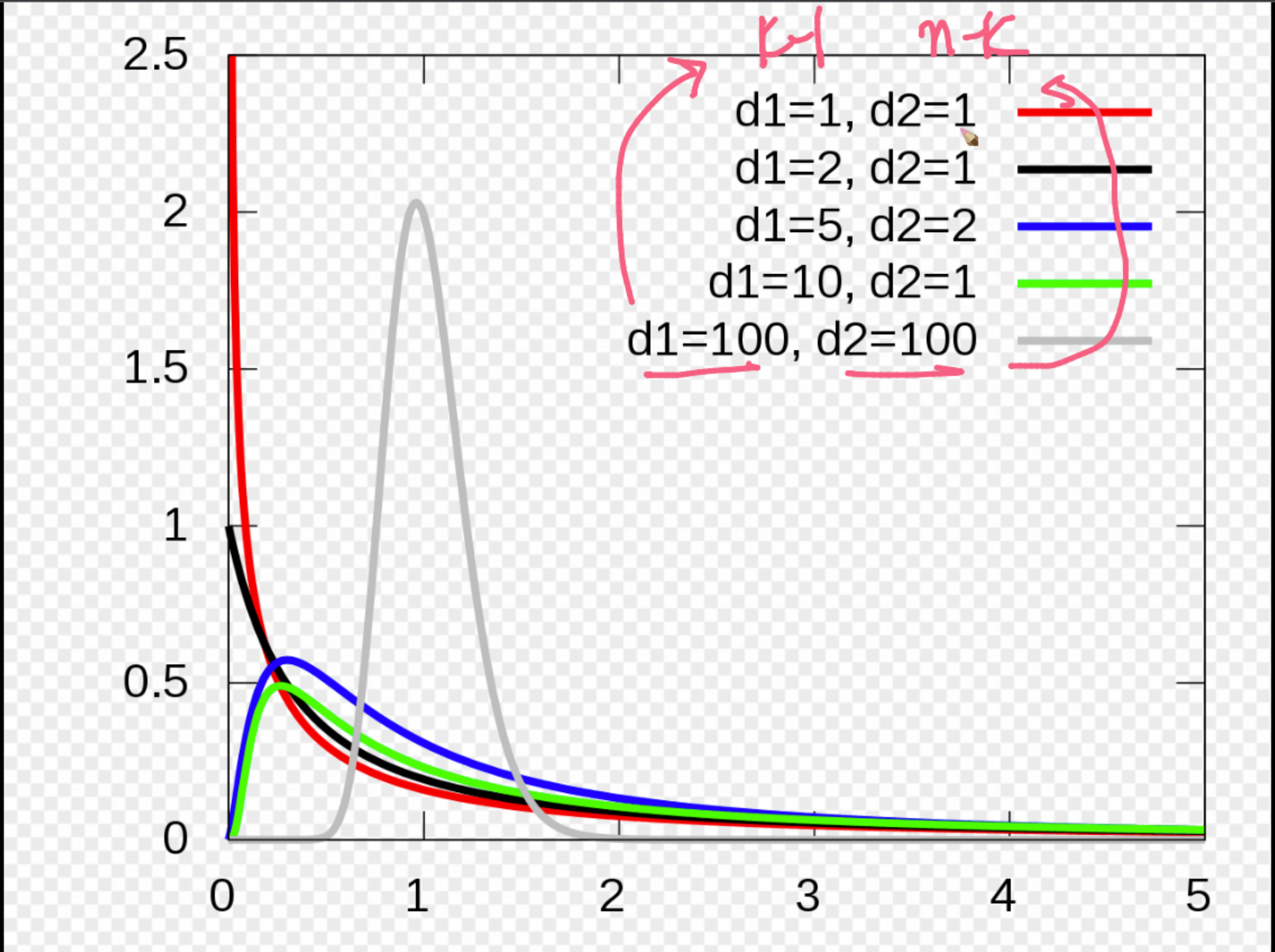
Makes sense ← $\underline{f}$ → +ve

dist under H_0

F-dist
$(k-1, n-k)$

$\boxed{H_0}$





The pdf of the F distribution.

p-val vs α

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:35

ANOVA: Assumptions:

1. each gp's observations are gaussian

   $G_1: 10 \quad G_2: 13$

2. each gp's variance is (practically) roughly the same

   $G_1: 10$ ✗
   $G_2: 1000$

3. each obs is independent ✓

Under H_0: Test stat (f) $\sim$ F-distb($k-1, n-k$)

33 / 39

Q

drug1 drug2 drug3

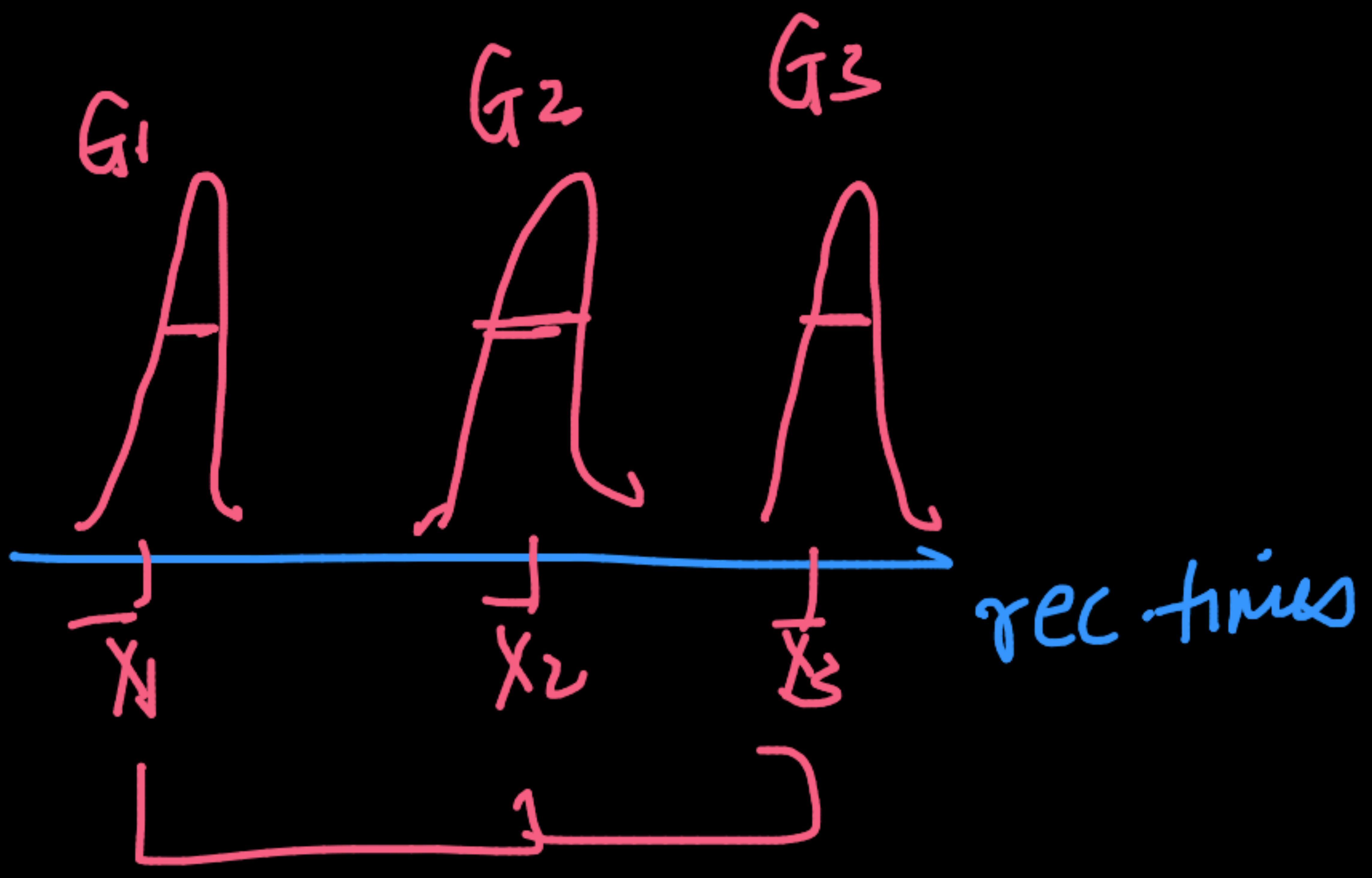
$\bar{X}_1 \quad \bar{X}_2 \quad \bar{X}_3 \quad \ldots$

are they gaussian?

NO

NO CLT

Q



$$f = \frac{MSB}{MSW}$$

$MSB < MSW$

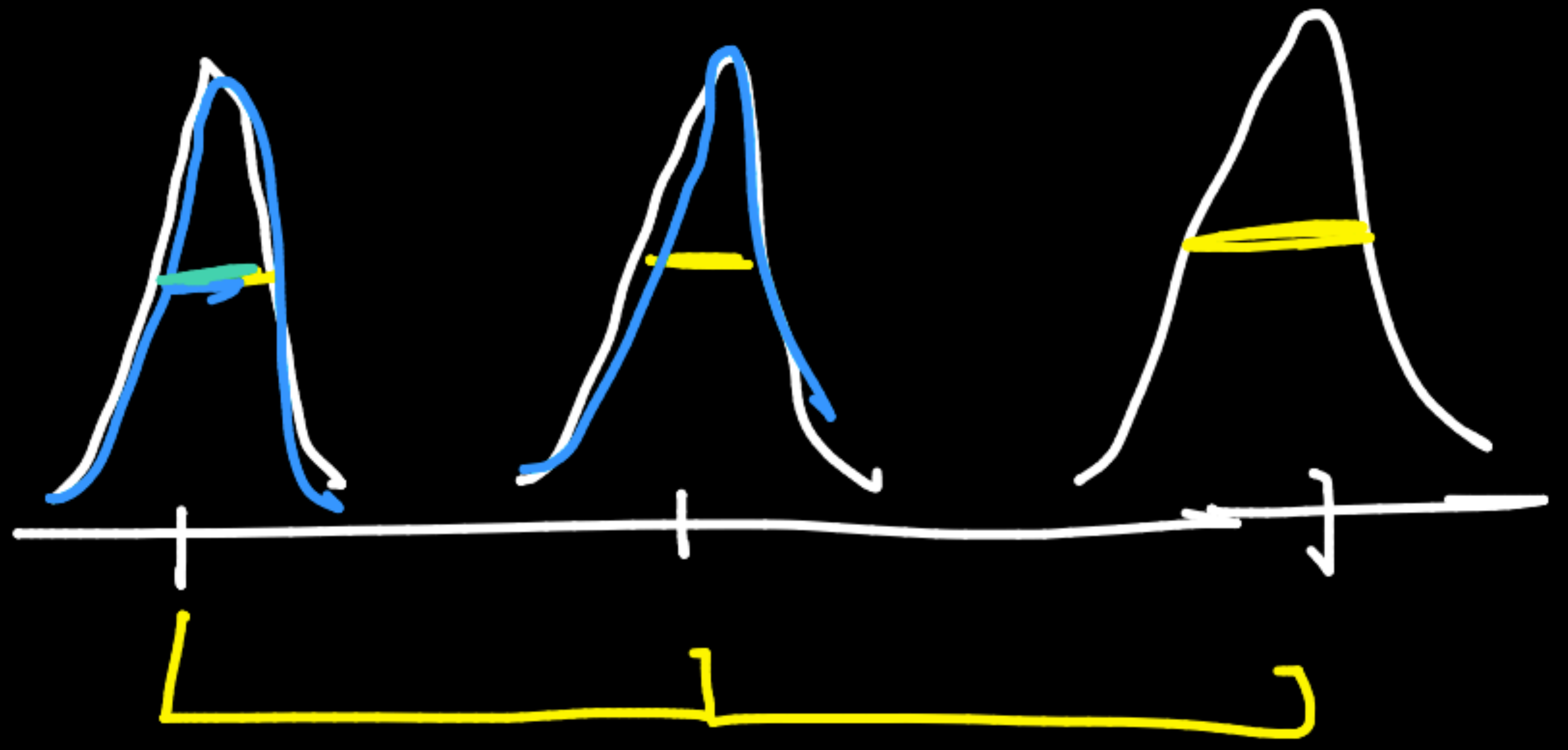
large

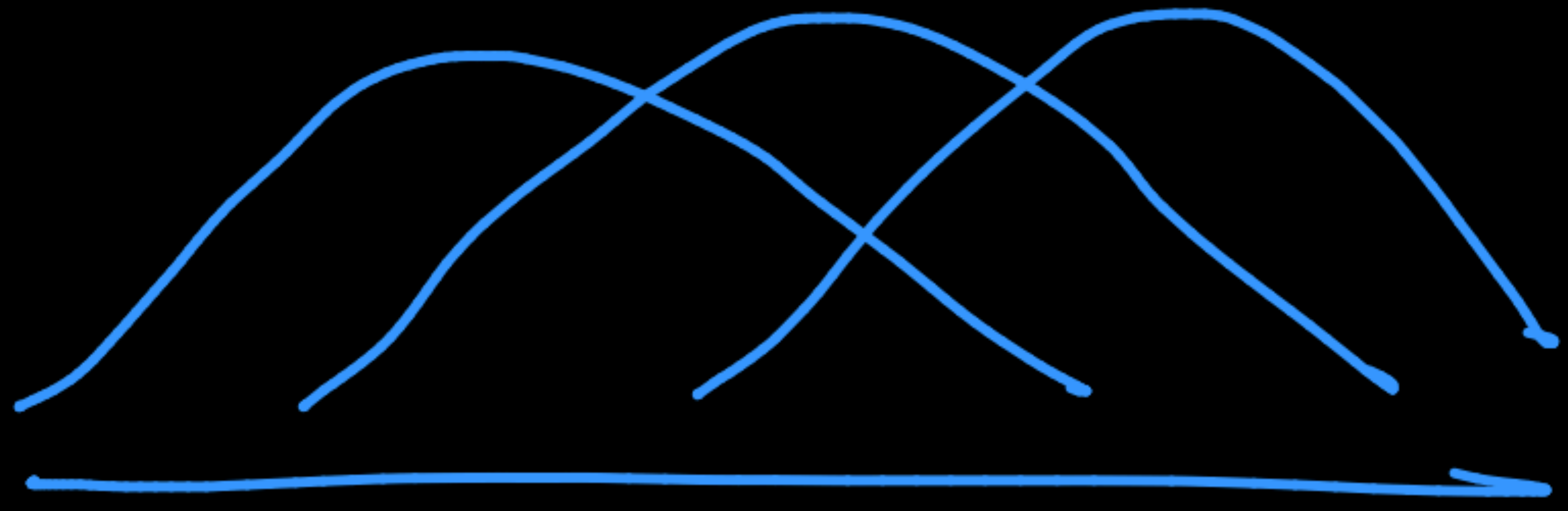
(small)

Q

accept Ha ✓

$$\uparrow f = \frac{MSB \uparrow}{MSW \downarrow}$$



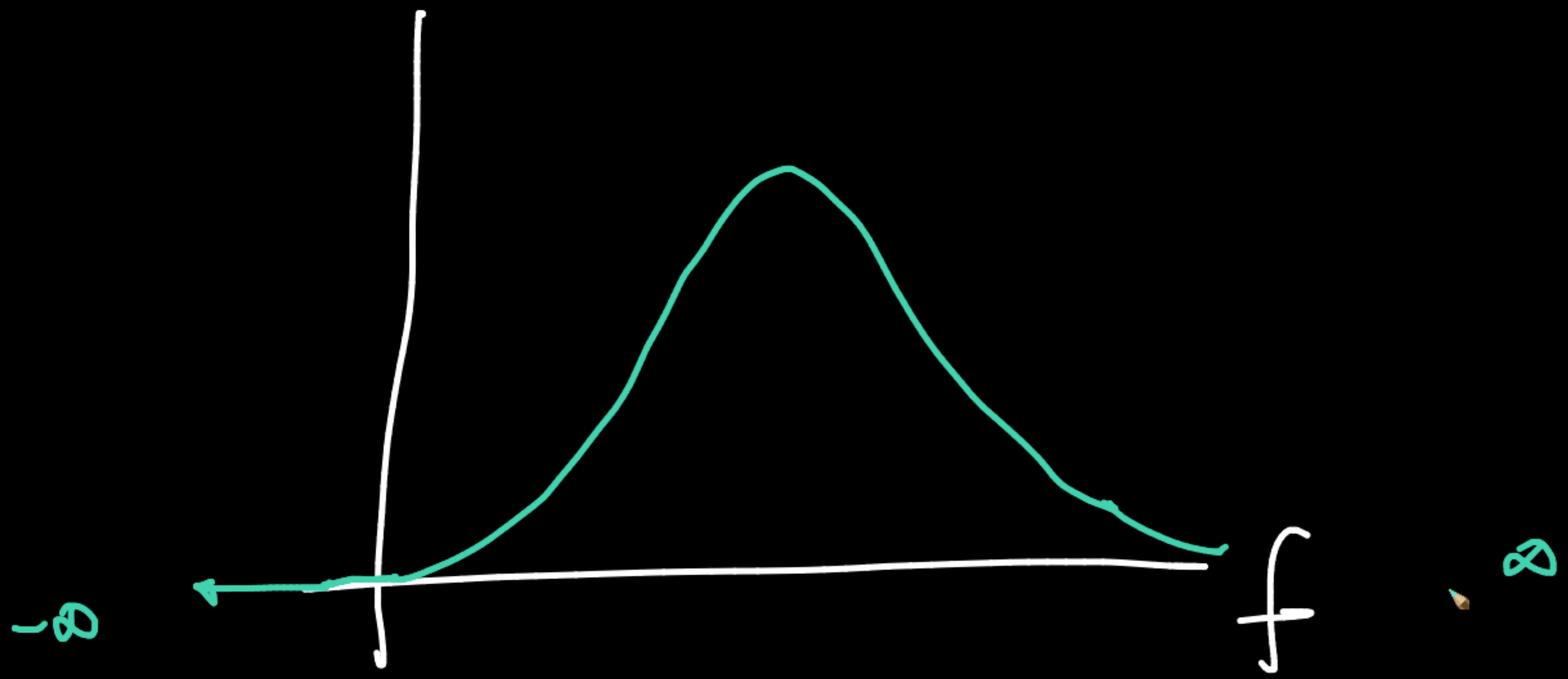


Q

Test-stat → abs ( ) or square

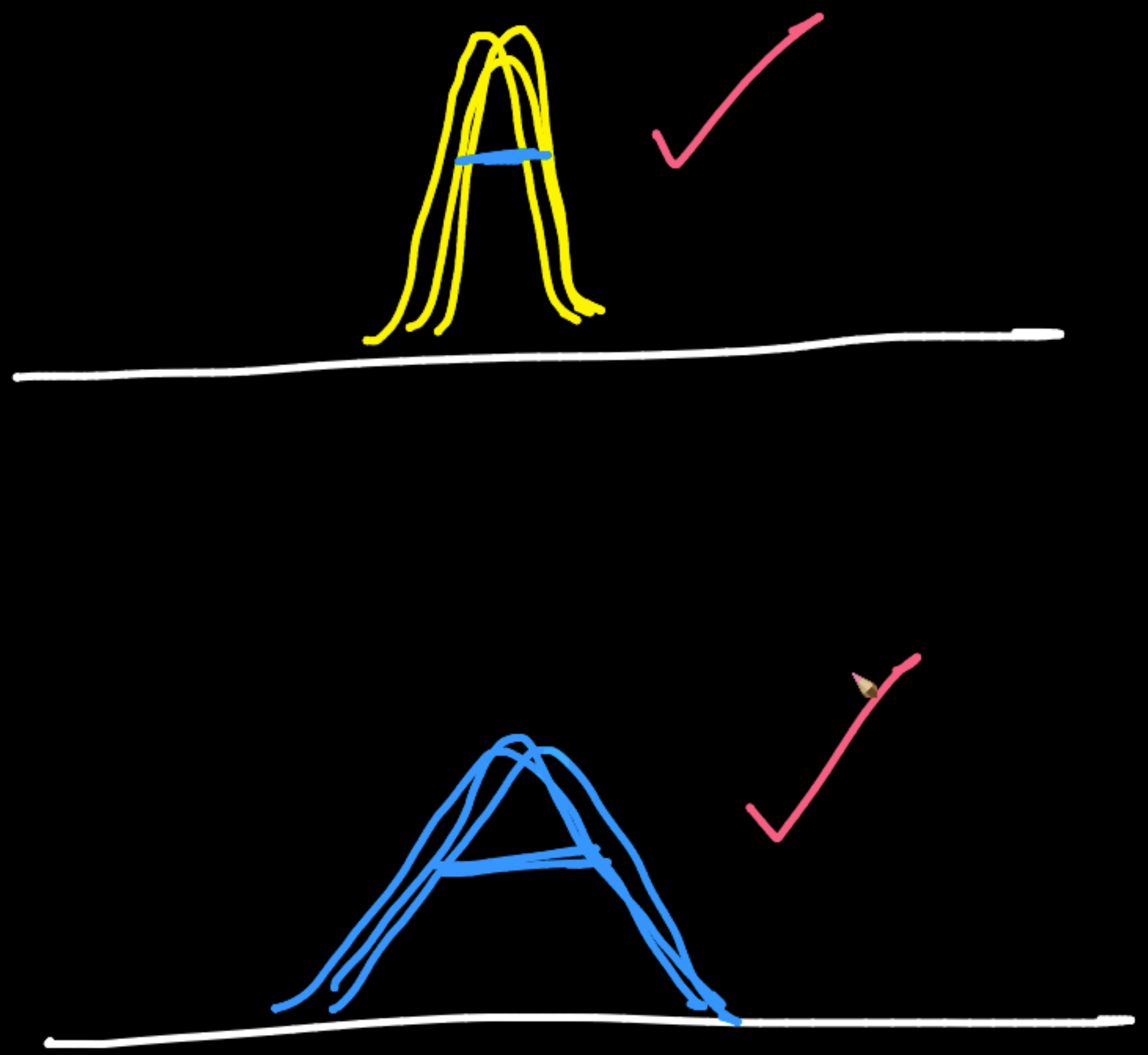
(> 0)

right-sided test (most)


-∞
f
∞

Q

Under Hb:



Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:37

Q

a

NBS

CLT / C.I

→ [2 3.5] ✓

[3 4.5]

α

p-value ✗

b

Hyp-test

$\boxed{\alpha}$: FP-error

40 / 40

ANOVA → 2+ grps

Adda lo

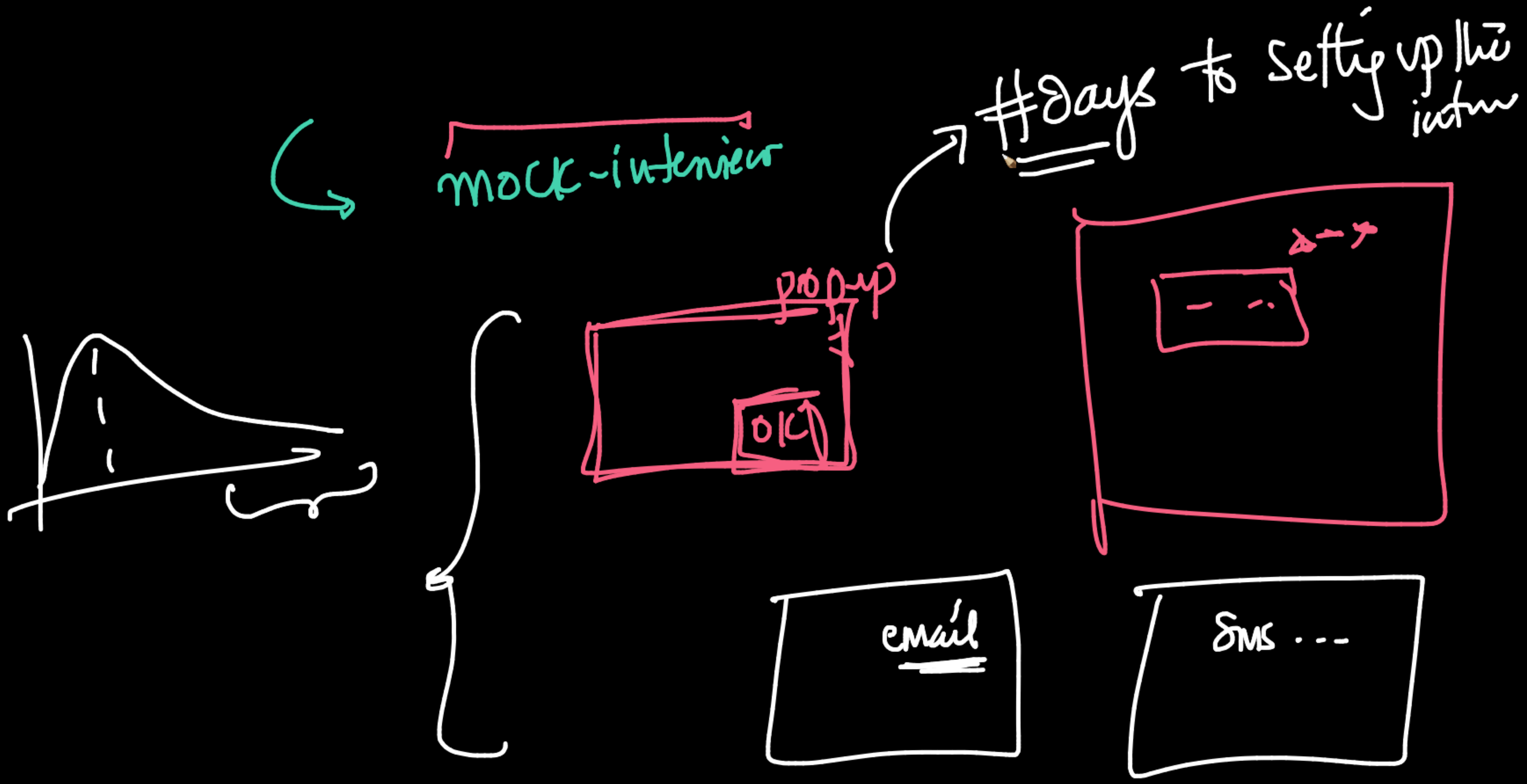
ecommerce:

B0 B1 B2 B3 . . . .

(mean sales in $)

th

C-1


mock-interview
#days to setup the interview
popup
OK
email
SMS ...

$\underline{\underline{\phi}}$

each gp is Gaussian

Variances of each gp are Same

Chrome File Edit View History Bookmarks Profiles Tab Window Help

Q

m samples per gp

extensions of ANOVA

$m_1 \quad m_2 \cdots \quad m_k$

$n = m_1 + m_2 + \cdots + m_k$

44 / 46

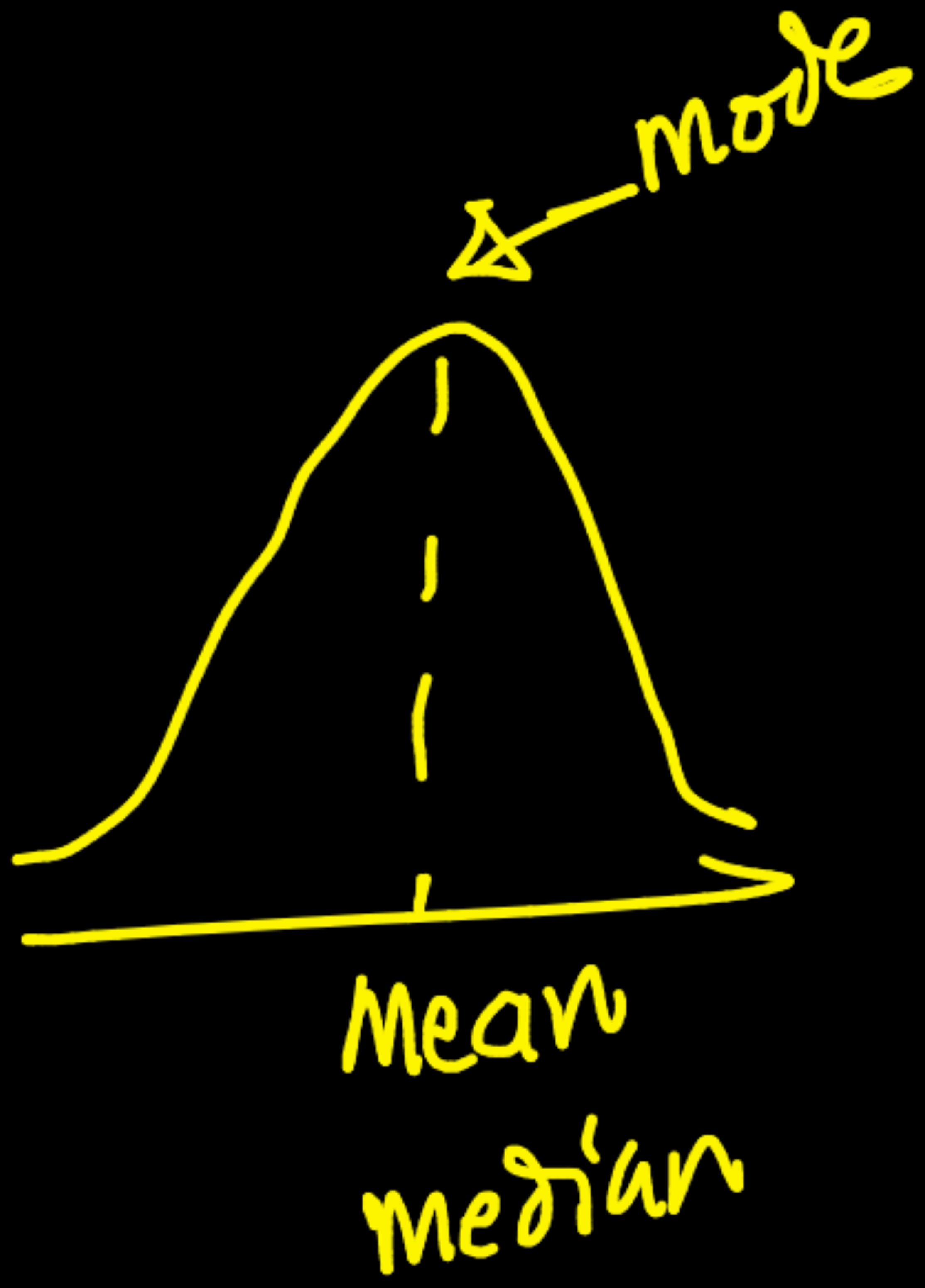
$\boxed{Q}$

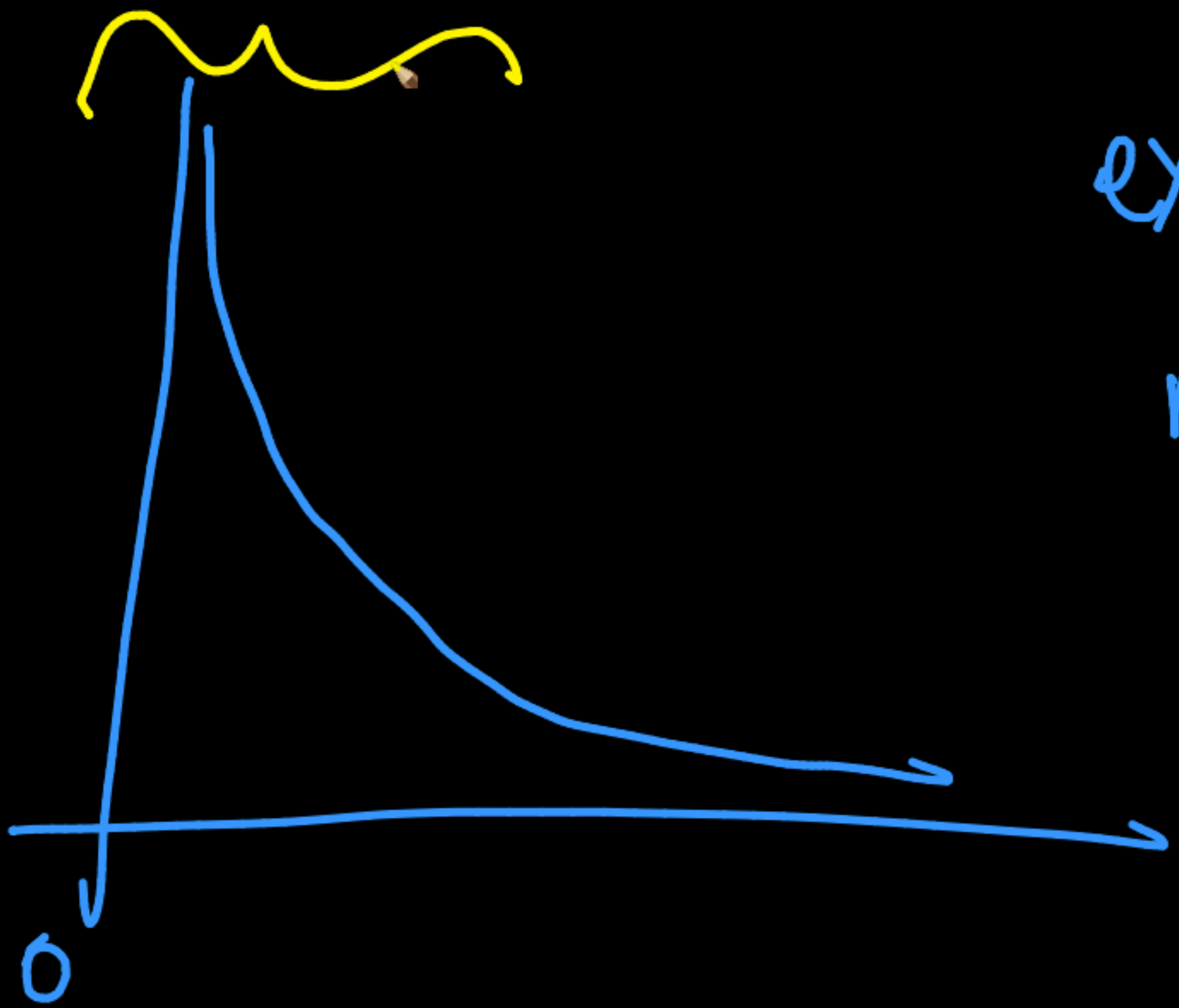
$$f = \frac{\text{variance of gp. sample means}}{\text{mean of within group variances}}$$

Q

Test-stat ≥ 0 $\Rightarrow$ non-param test

ANOVA





$\text{expo}(\lambda)$

$\text{mean-expo} = 1/\lambda$

peak = mode may|mayn't be the mean

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:49

Q

$\boxed{\text{Test-stat} = f}$

$\boxed{H_0}$

dist of test-statistic under H_0

$\downarrow$

F-dist$(k-1, n-k)$

48 / 48

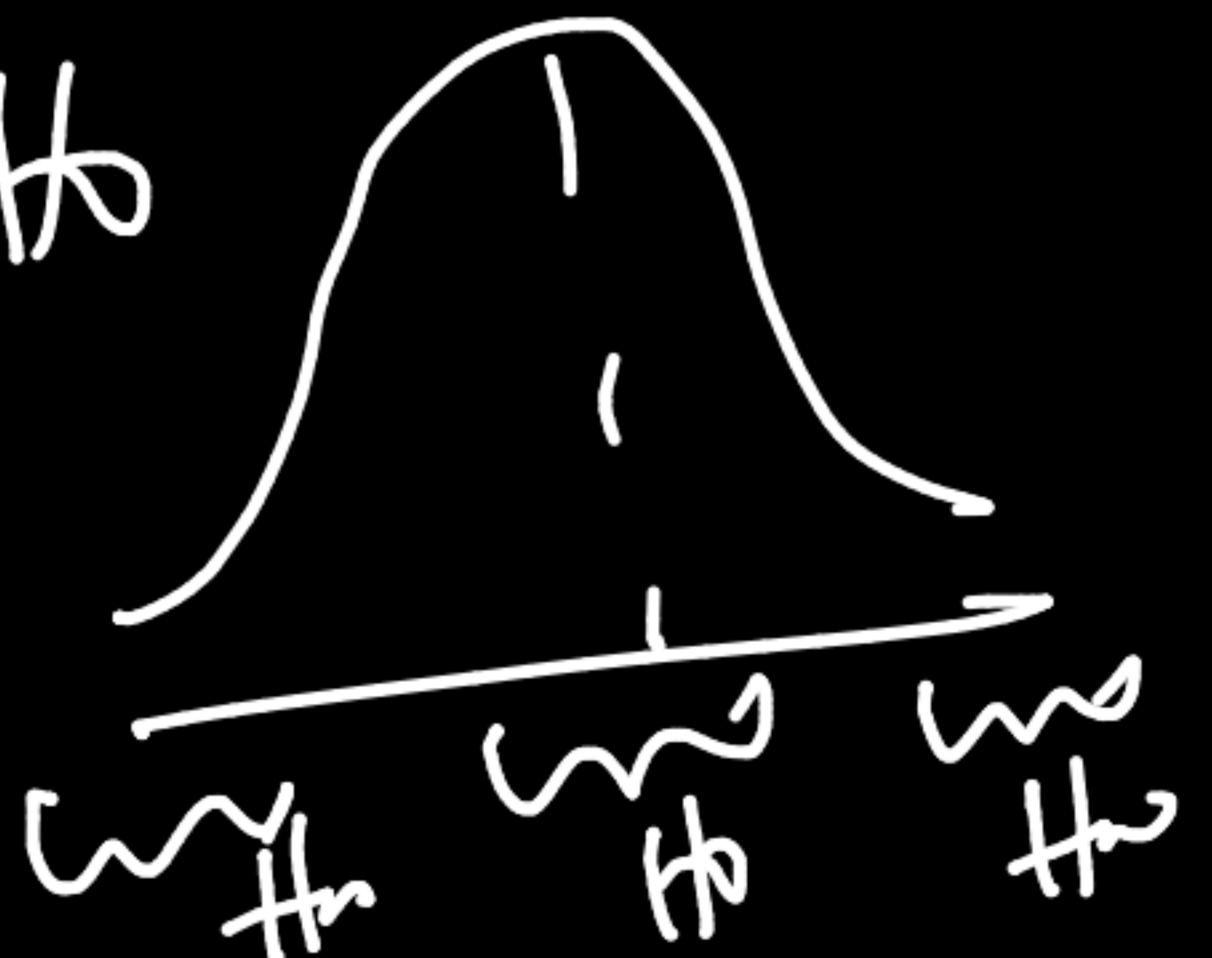
$\boxed{\text{Z test:}}$ ✓

$$T = \frac{\bar{X}_1 - \bar{X}_2 - \cancel{(\mu_1 - \mu_2)}}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}$$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

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

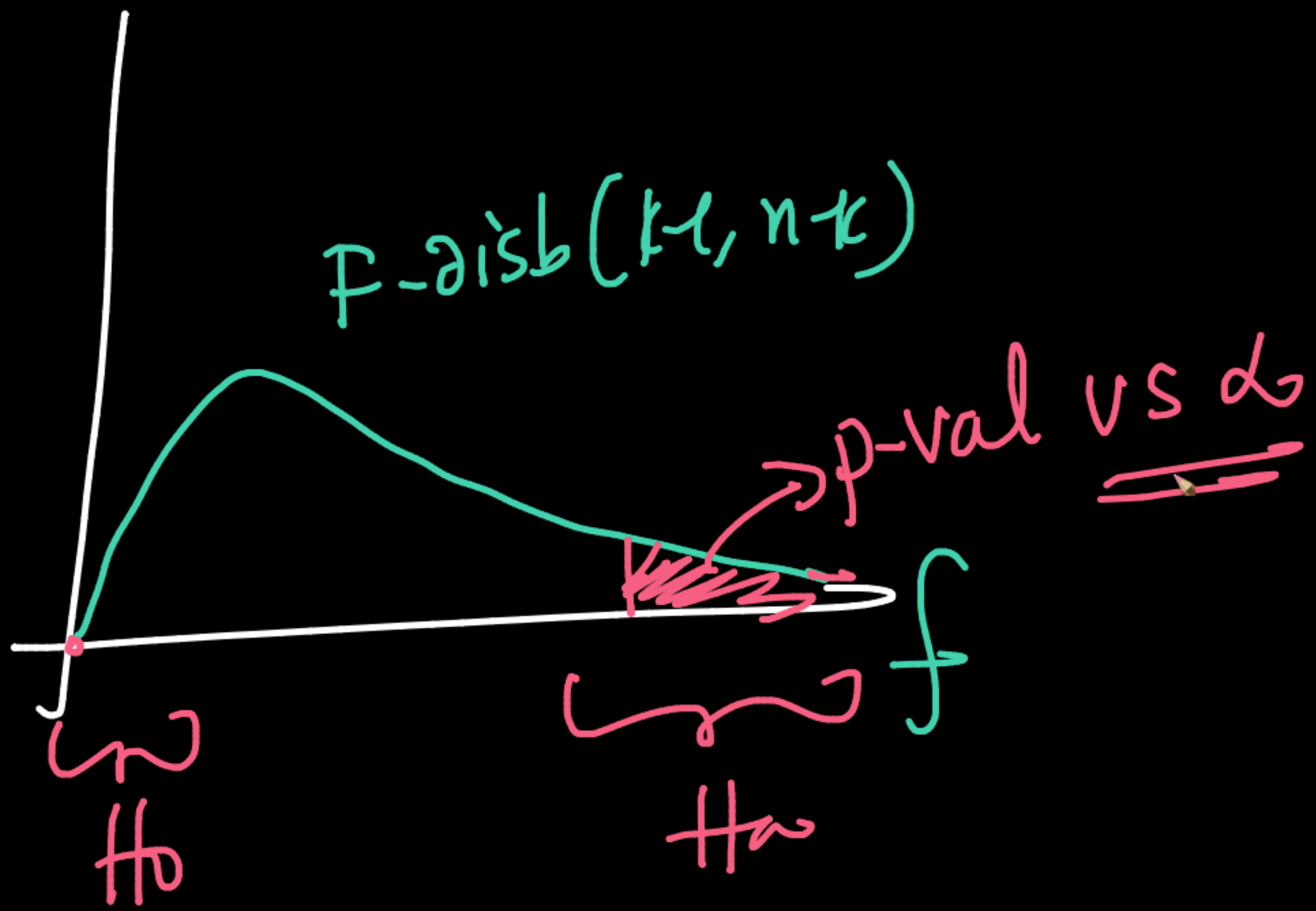


Q

ANOVA: Comparison means of

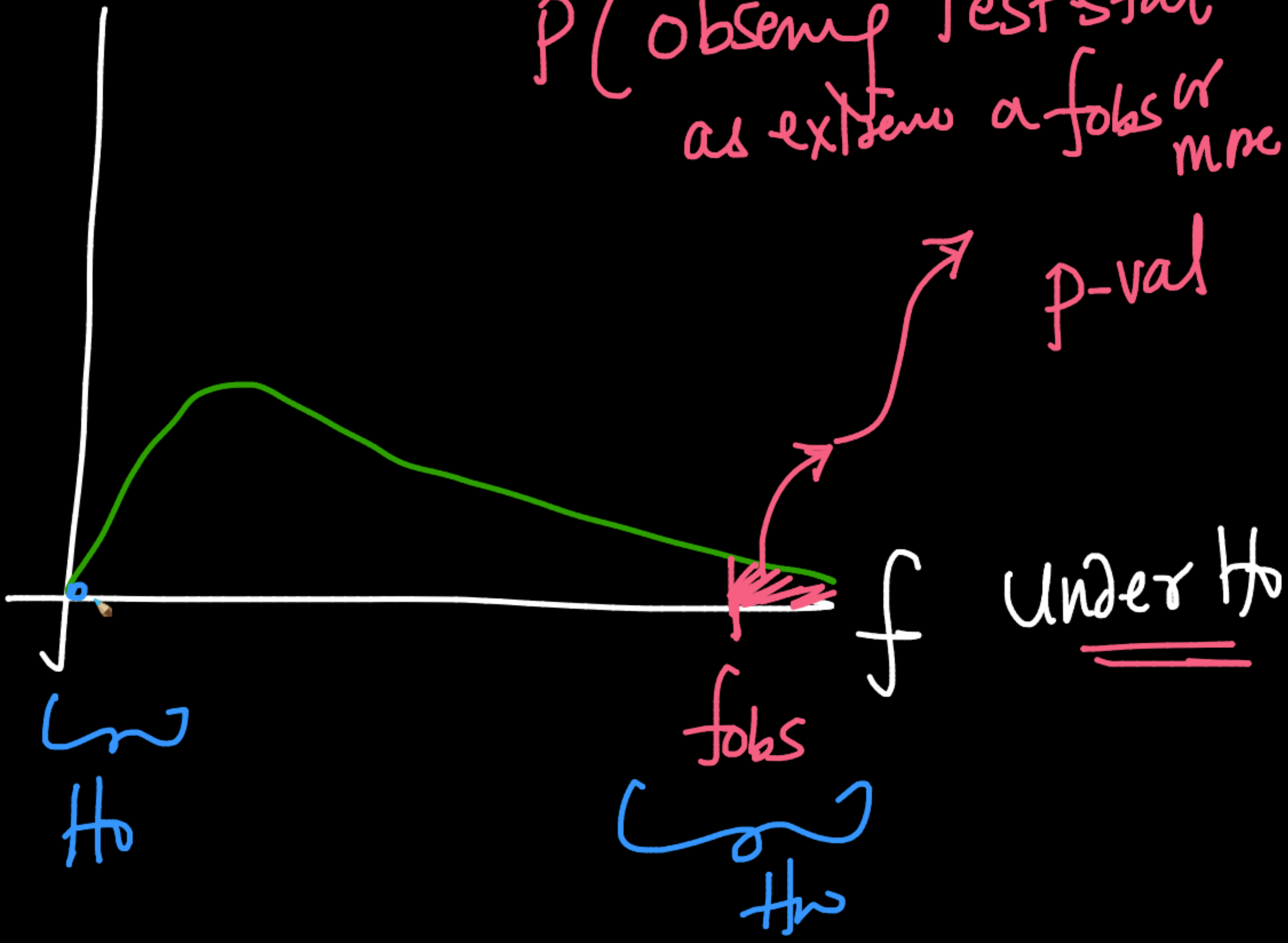
$k \geq 2$ gps

Q



P(observing Test stat as extreme as fobs or more | Ho)

p-val



ANOVA, Variations

— Types of variations

one dependent var

#days to recover

Oneway

d_1, d_2, d_3

indep-var

2-way

drug: d_1, d_2, d_3

age: K, A, O

8

2-indep-var

Multi-way ANOVA

d_1, d_2, d_3 ← drug

K, A, O ← age-gp

SA; SEA; AA : Cau ← ethn

multiple indep var

(features)

Stats

{Multi-way ANOVA} ⟷ [Linear Regr (ML)]

Special Case

[Geom & Linear Algebra]

MANOVA:

one-way Manova

dep.var

Salary

IQ

College-degree

[HS, B, M, PhD]

indep-var

(one)

2-way Manova

dep. var

College-degree
(HS/B/M/PhD)

Salary

2 indep-
var

IQ

Ethnicity
(SA, SEA, Cau, AA ..)

2 or more
dep. var

Any HT or

Variation of ANOVA } → framework & conceptual

Chrome File Edit View History Bookmarks Profiles Tab Window Help

p-hacking

Multiple / too many HT → Same α = 5%.

α =

$^{k}C_2$ test

{ high prob of making mistakes }

solns:

↳ ANOVA like methods

use C.I

→ Stats :- Adjusted p-value

58 / 60

$\alpha = 5\%$

$\overbrace{50 \quad 50 \quad 50}$

$d_1 \quad d_2 \quad d_3$

$\mu_1 \neq \mu_2$

Data

100 obs ← D1

100 obs ← D2

100 obs ← D3

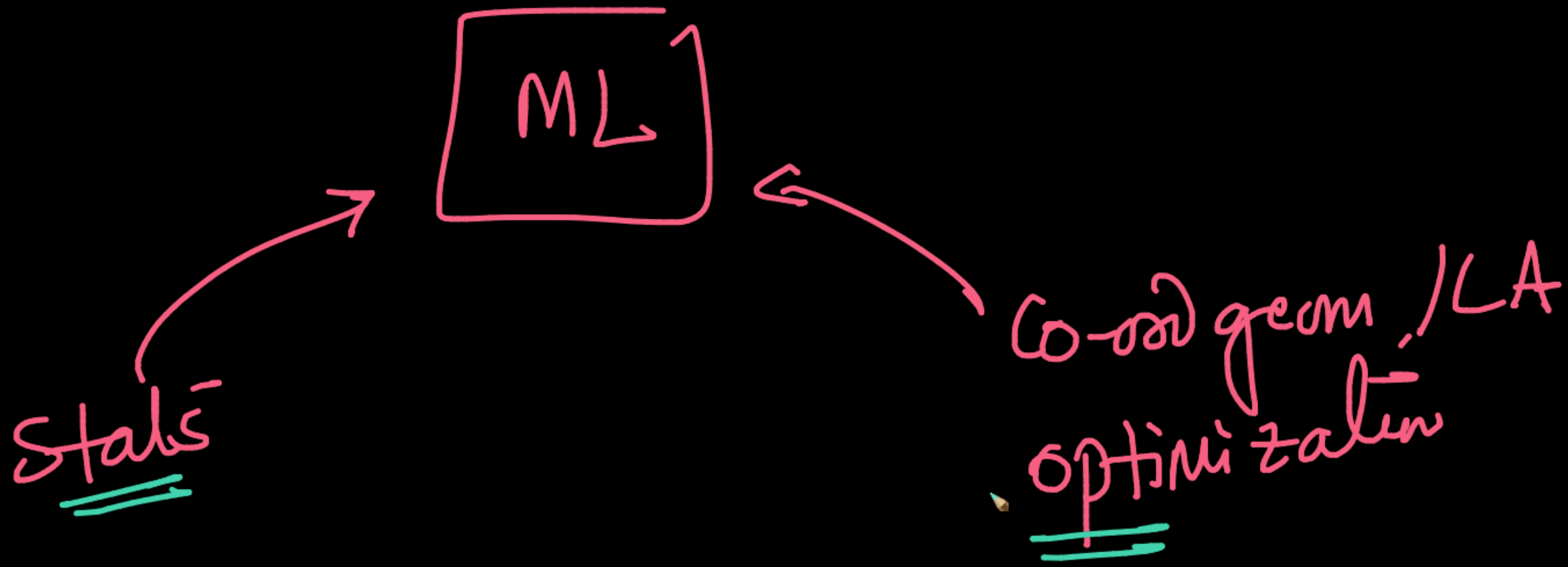
Unseen

50 d_1

50 d_2

50 d_3

ML


ML
Stats
Co-ord geom /LA
optimization

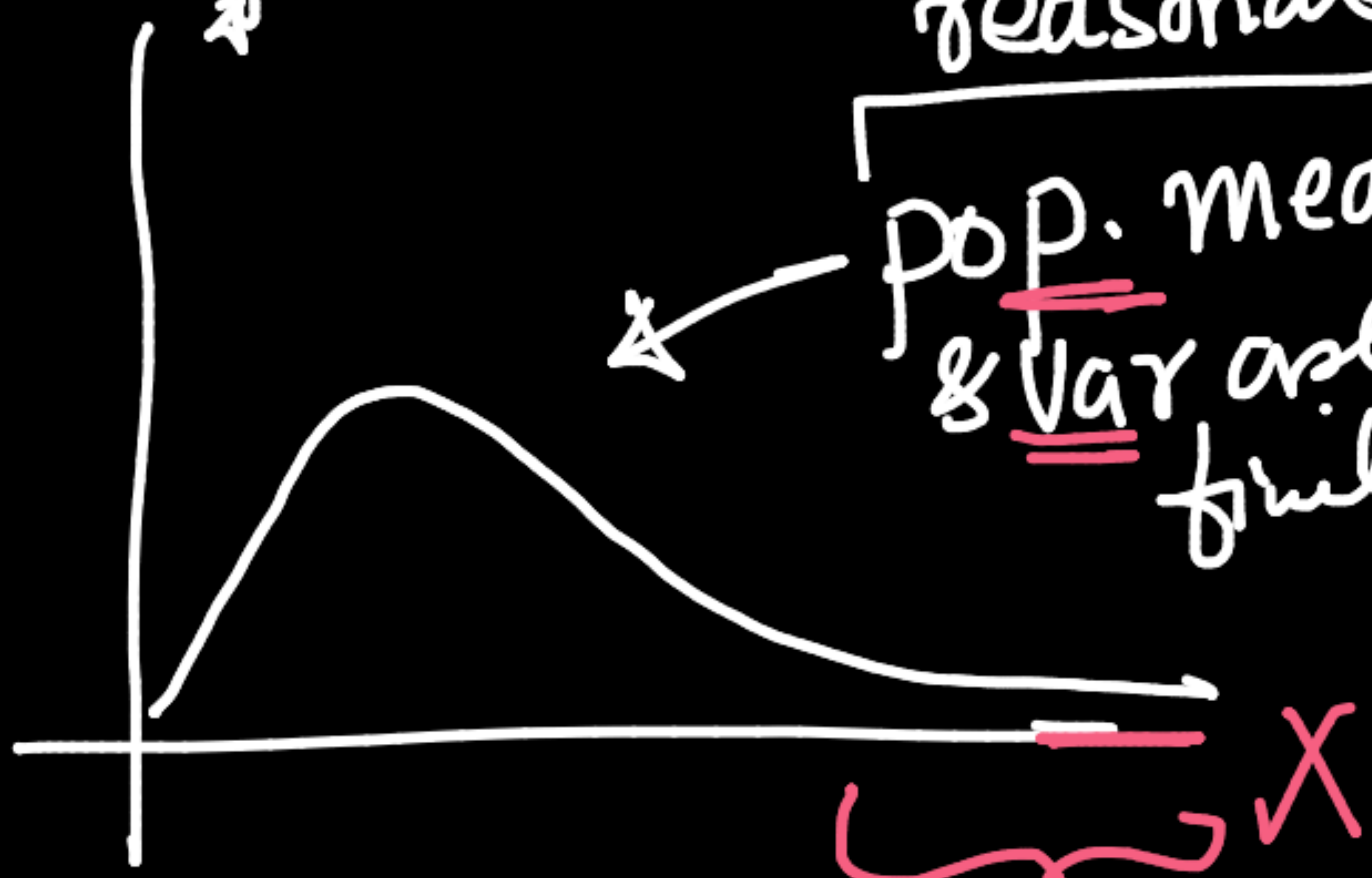
Q

$X : x_1, x_2 \ldots x_n < 30$

t-test ✓

✓

reasonable pop. mean & var are finite



log-normal → right skew

Pareto dist ($\alpha = 1$)

↳ μ & $\sigma = \infty$

Q

Compare means

1 → 1 Sample Test

2 → 2 Sample Test

>2 → ANOVA

Pop. mean & σ are finite

n_1 & $n_2 > 30$
and σ_1 & σ_2 → z-test

else → t-test

✓ Compare prop $\rightarrow$ Z-prop. tests

n_1 & $n_2 \geq 30$

Test of indep

$\boxed{\text{Cat}}$ $\rightarrow$ Gender $\leftarrow$ Purchase } $\rightarrow$ χ^2-test

W H.

numerical ;- Weight & height

Corr. Coeff

Test of significance

Q

G \ Pre	0	1
F	5	4
M	3	2

–	–
–	.

obs

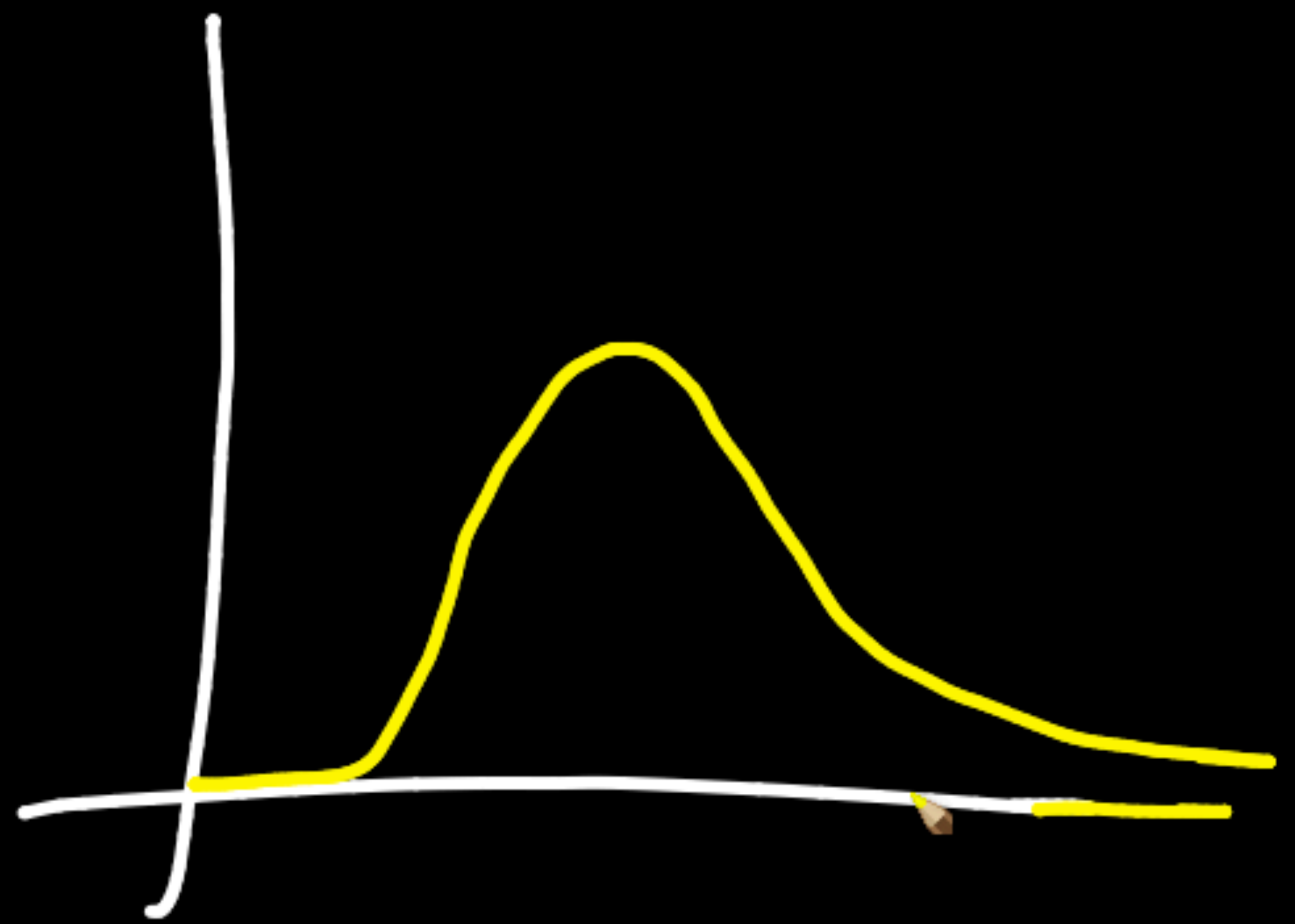
less likely to reject H_0

p-val

Q

KS-test

$$T = \sup |\text{gap b/w dist}|$$



$$\sqrt{n}\,T \sim \text{Kolmogorov dist} \quad \text{as } n \to \infty$$

Physics

KS-test: $\begin{cases} X \sim \boxed{N}(\mu, \sigma_1) \\ \quad\quad\quad\;\; \| \quad\; \| \\ Y \sim \boxed{N}(\mu, \sigma_2) \end{cases}$

(NO)

accept H_0

m&s

KS-test

$X: \underbrace{x_1, x_2 \ldots x_n}$

$N(m, s) \xrightarrow{\text{r.vs}} \underbrace{x_1' \; x_2' \; x_3' \ldots x_n'}_{\text{m\&s}}$

Q

FP-error $= \underline{\underline{\alpha}}$ (sig. level)

$\underline{\underline{\beta}}$: FN $\underline{\underline{\text{error}}}$

power $= \underline{\underline{1-\beta}}$

→ Test

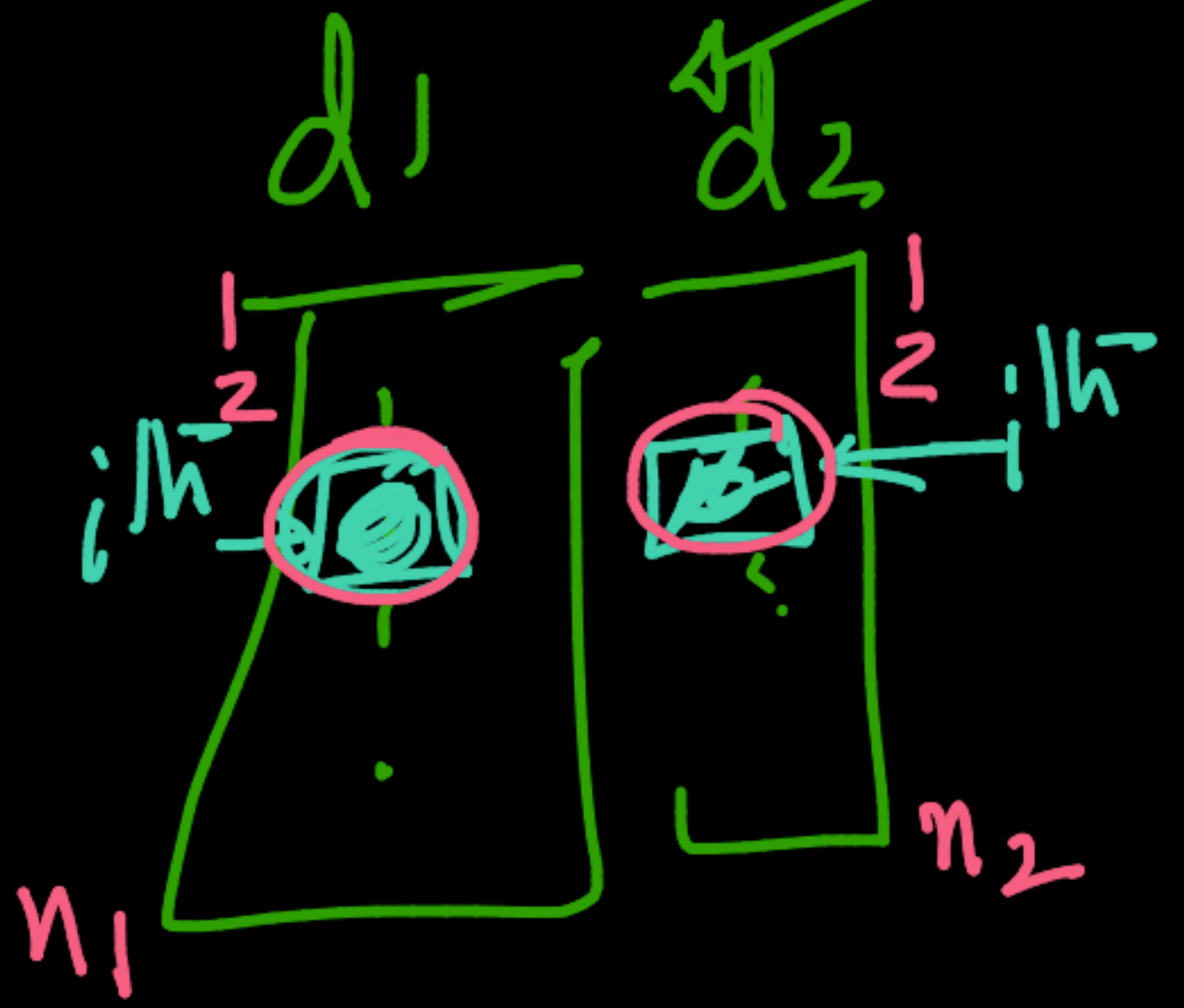
→ 'n'

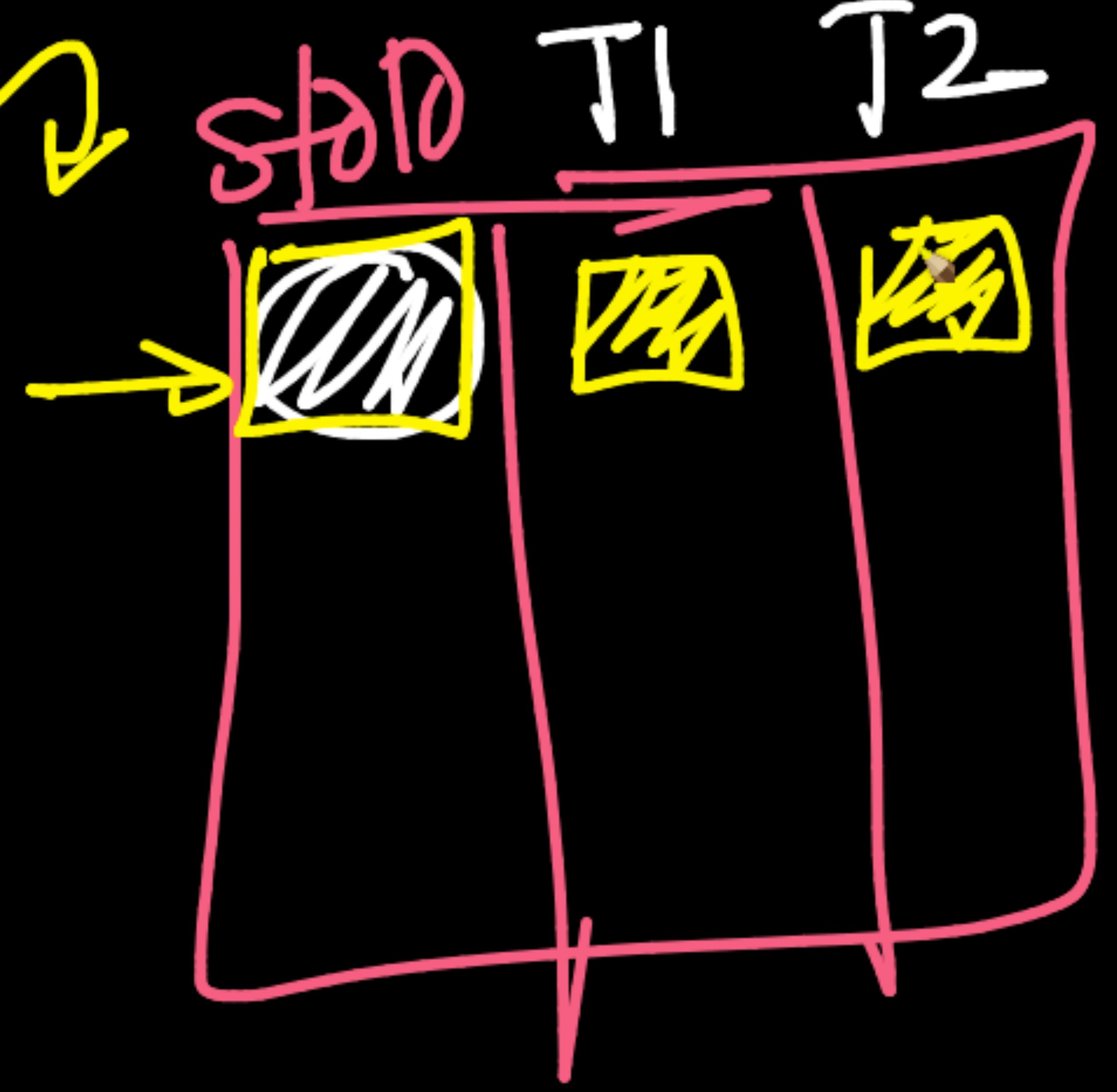
. . .

Q

independent t-test

dep t-test





Q

p-hacking

adj. p-value

lower α for each test

Q

pop μ & σ are finite

$\boxed{\text{CLT}}$

T-test

data need not be normal

$x_1, x_2 \ldots x_n$

$\bar{X} \sim$ Normal

pop μ & $\sigma \neq \infty$

Assumptions [edit]

Most test statistics have the form $t = \frac{Z}{s}$, where Z and s are functions of the data.

Z may be sensitive to the alternative hypothesis (i.e., its magnitude tends to be larger when the alternative hypothesis is true), whereas s is a scaling parameter that allows the distribution of t to be determined.

As an example, in the one-sample t-test

Sample- mean

$$t = \frac{Z}{s} = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}}$$

where $\overline{X}$ is the sample mean from a sample $X_1, X_2, \ldots, X_n$, of size n, s is the standard error of the mean, $\widehat{\sigma}$ is the estimate of the standard deviation of the population, and μ is the population mean.

The assumptions underlying a t-test in the simplest form above are that:

- $\overline{X}$ follows a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$
- $s^2(n-1)/\sigma^2$ follows a χ^2 distribution with $n-1$ degrees of freedom. This assumption is met when the observations used for estimating s^2 come from a normal distribution (and i.i.d for each group).
- Z and s are independent.

In the t-test comparing the means of two independent samples, the following assumptions should be met:

- The means of the two populations being compared should follow normal distributions. Under weak assumptions, this follows in large samples from the central limit theorem, even when the distribution of observations in each group is non-normal.[18]
- If using Student's original definition of the t-test, the two populations being compared should have the same variance (testable using F-test, Levene's test, Bartlett's test, or the Brown–Forsythe test; or assessable graphically using a Q–Q plot). If the sample sizes in the two groups being compared are equal, Student's original t-test is highly robust to the presence of unequal variances.[19] Welch's t-test is insensitive to equality of the variances regardless of whether the sample sizes are similar.
- The data used to carry out the test should either be sampled independently from the two populations being compared or be fully paired. This is in general not testable from the data, but if the data are known to be dependent (e.g. paired by test design), a dependent test has to be applied. For partially paired data, the classical independent t-tests may give … endent t-test is sub-optimal as it discards the unpaired data.[20]

Misc

bias of an estimate

$\frac{1}{n-1}$ den fr var

Chrome File Edit View History Bookmarks Profiles Tab Window Help 12:01

Q

Bernouli (p)

$\alpha = 5\%$

$H_0: p = 0.6$

mean : p

std-dev: p(1-p)

1 0 1 0 0 0 - -

95C.I

[,]

76 / 77

Q

ANOVA

$H_0: \mu_1 = \mu_2 = \mu_3$

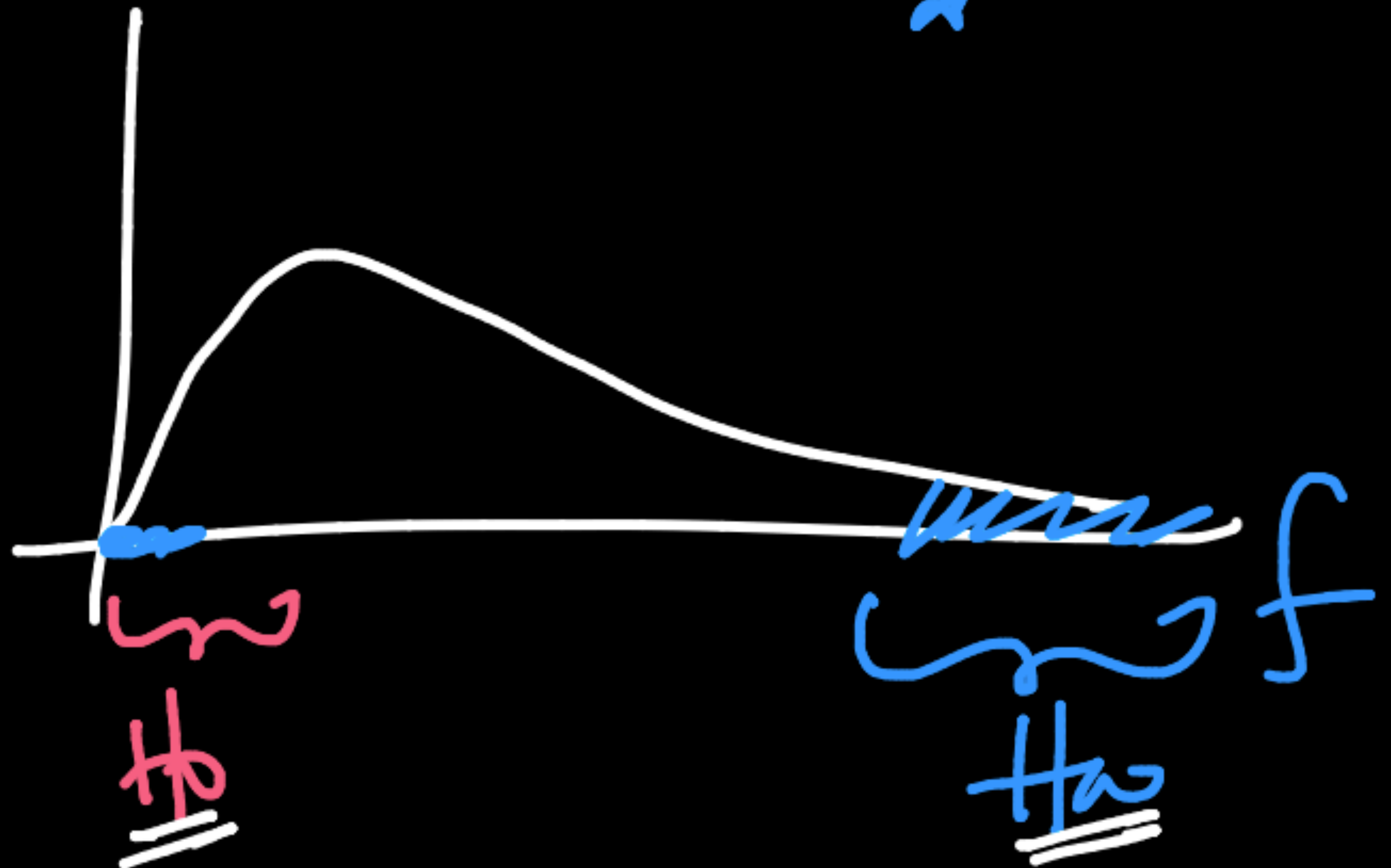
$H_a: \boxed{\mu_1 < \mu_2 < \mu_3}$

all pop means are not same

$\mu_1 \neq \mu_2 = \mu_3$

$\mu_1 = \mu_2 \neq \mu_3$

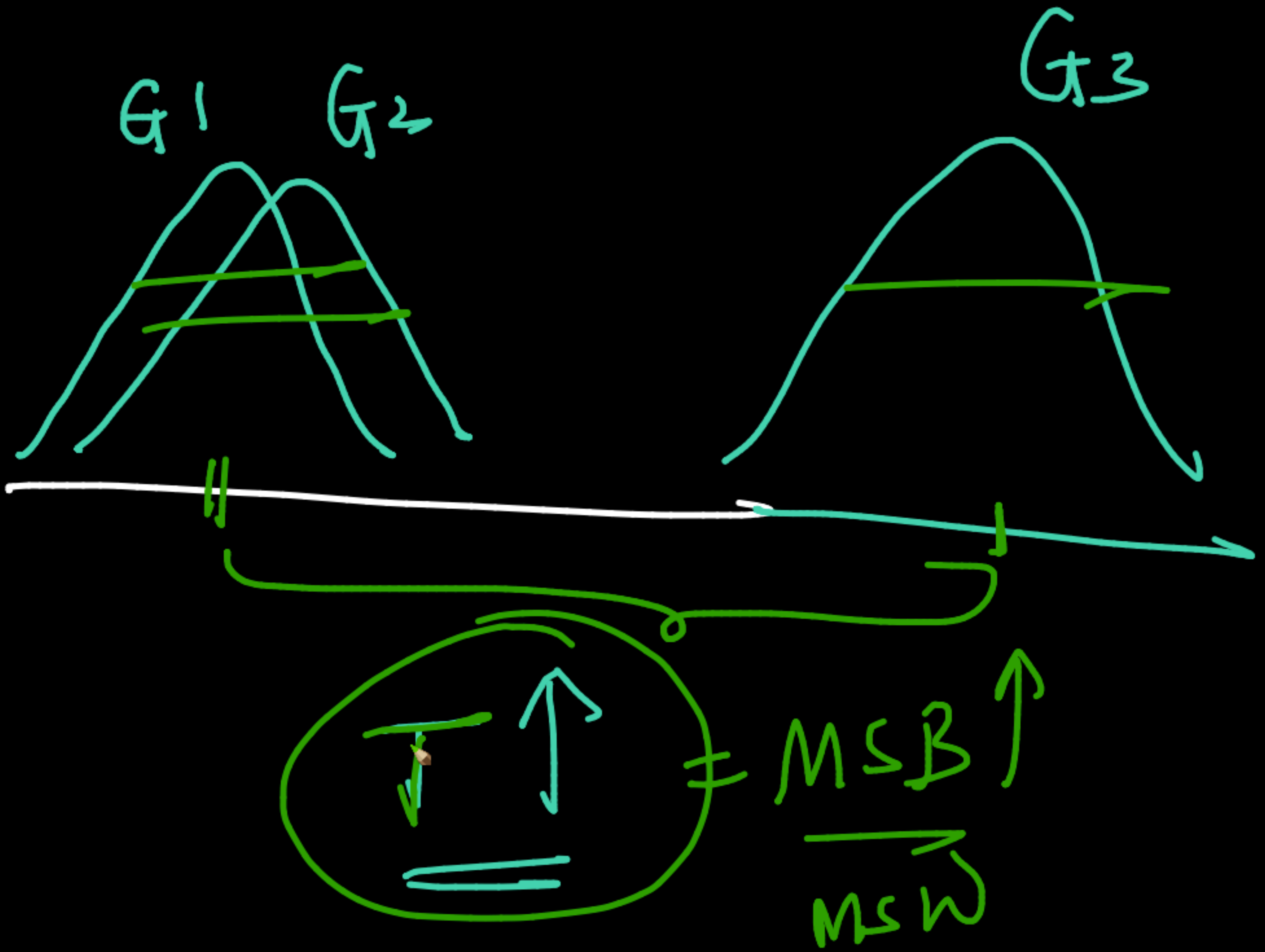
left tailed test



$F = \dfrac{MSB}{MSW}$

$$\mu_1 \approx \mu_2 < \mu_3$$

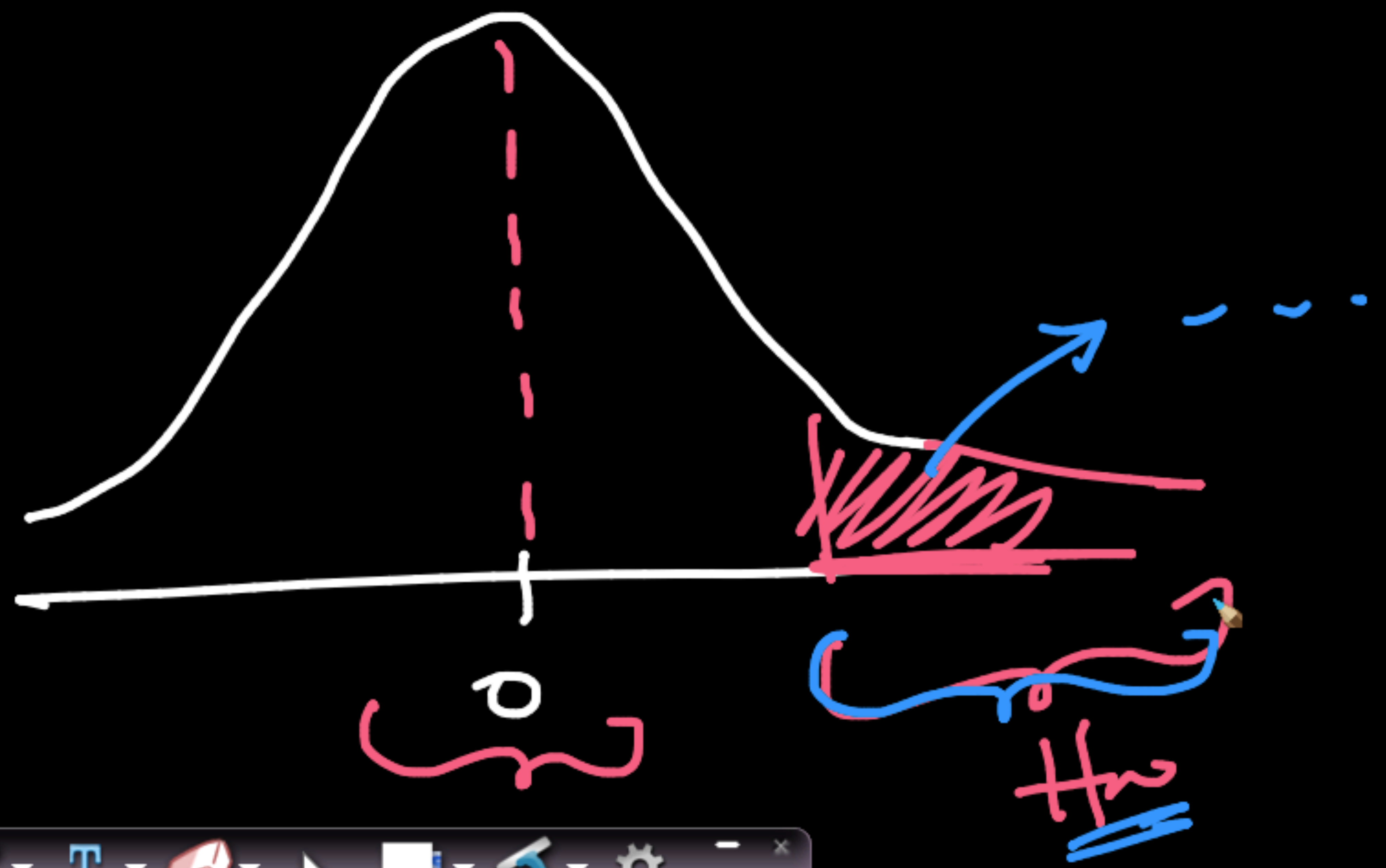
$$\mu_1 < \mu_2 \approx \mu_3$$



$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2$

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



Anou A:

Q

KS-Test

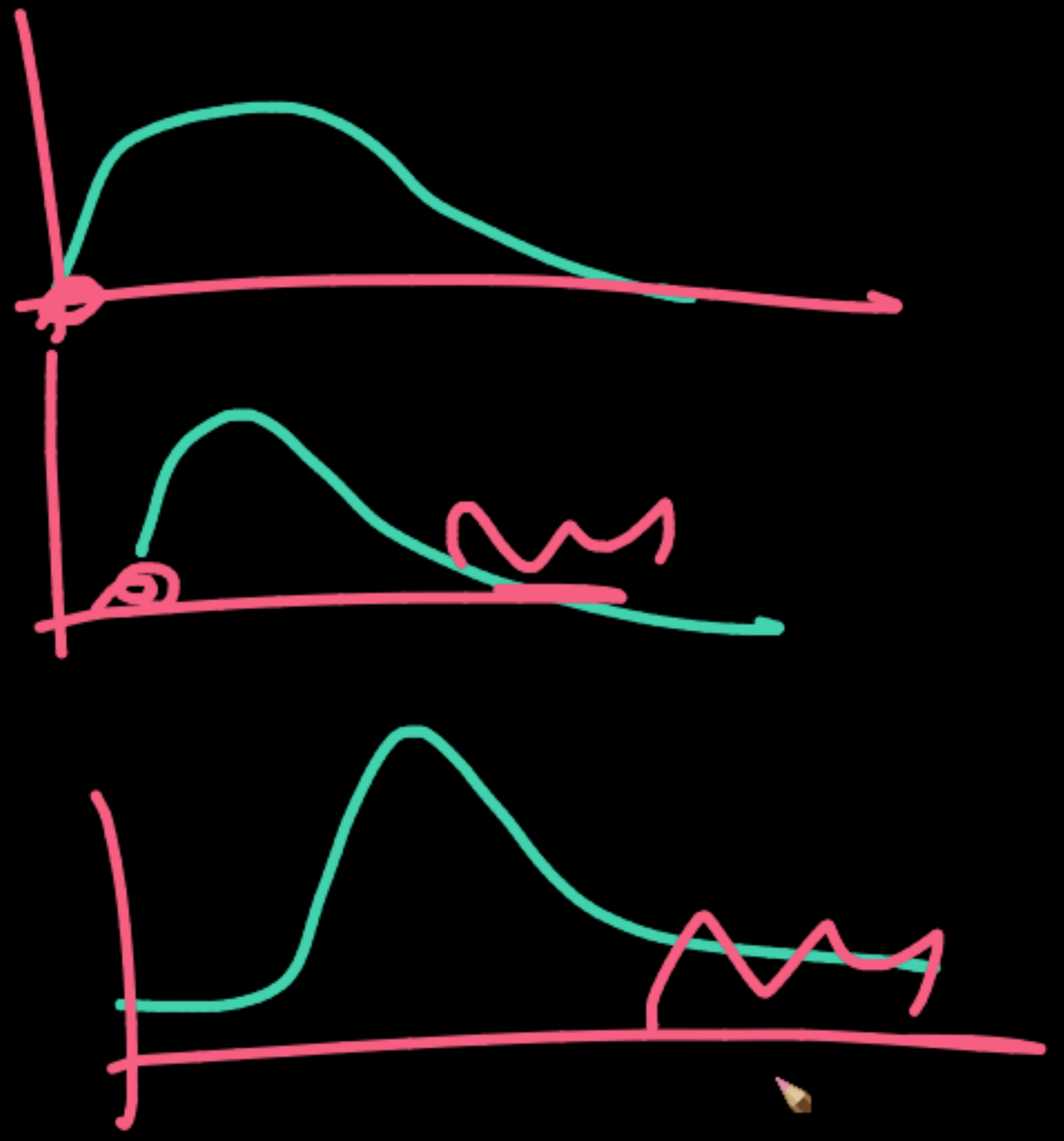
$$\left.\begin{array}{l} X \sim N(\mu, \sigma) \\ Y \sim N(k\mu, k\sigma) \end{array}\right] \rightarrow \text{Not the same} \ldots$$

Q

$$\chi^2 \rightarrow T \geq 0$$

$$\text{ANOVA} \rightarrow T \geq 0$$

$$\text{KS-test} \rightarrow T \geq 0$$



right tailed tests

95% C.I of acc. true

d_1 — k → + var_1

d_1 A → + var_2

Q

Z-Test

$H_0: \mu_1 = \mu_2$ } ← Comparision of mean

$$\text{power} = 1-\beta$$ ✓

n, α, Test (z-test)

Control over

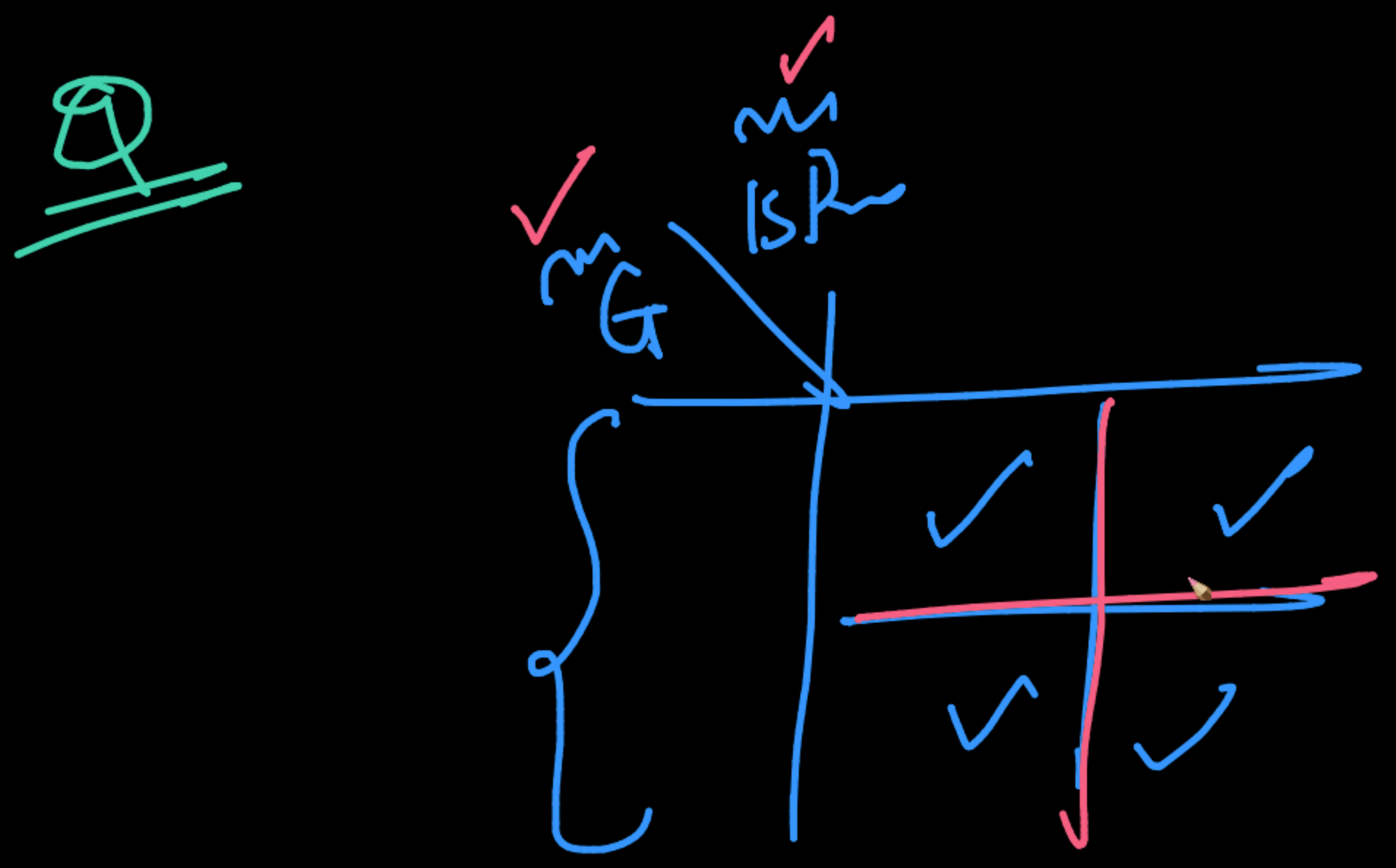
Test of Normality

$x_1 \dots x_n \sim$ Normal or not

KS-test

AD-test

more powerful


G
ISR

X_1, X_2

If array_like, it should be a 1-D array of observations of random variables, and the two-sample test is performed (and rvs must be array_like). If a callable, that callable is used to calculate the cdf. If a string, it should be the name of a distribution in `scipy.stats`, which will be used as the cdf function.

args : *tuple, sequence, optional*

Distribution parameters, used if *rvs* or *cdf* are strings or callables.

H_0: same dist
H_a: not same dist

N : *int, optional*

Sample size if *rvs* is string or callable. Default is 20.

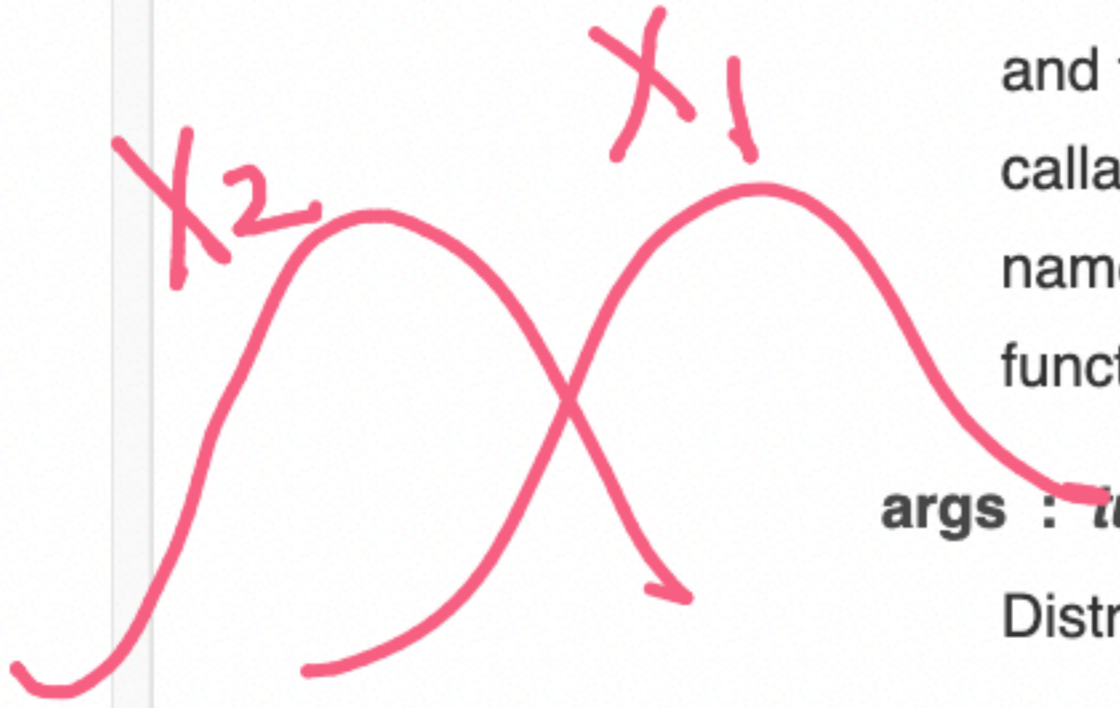
alternative : *{'two-sided', 'less', 'greater'}, optional*

Defines the null and alternative hypotheses. Default is 'two-sided'. Please see explanations in the Notes below.

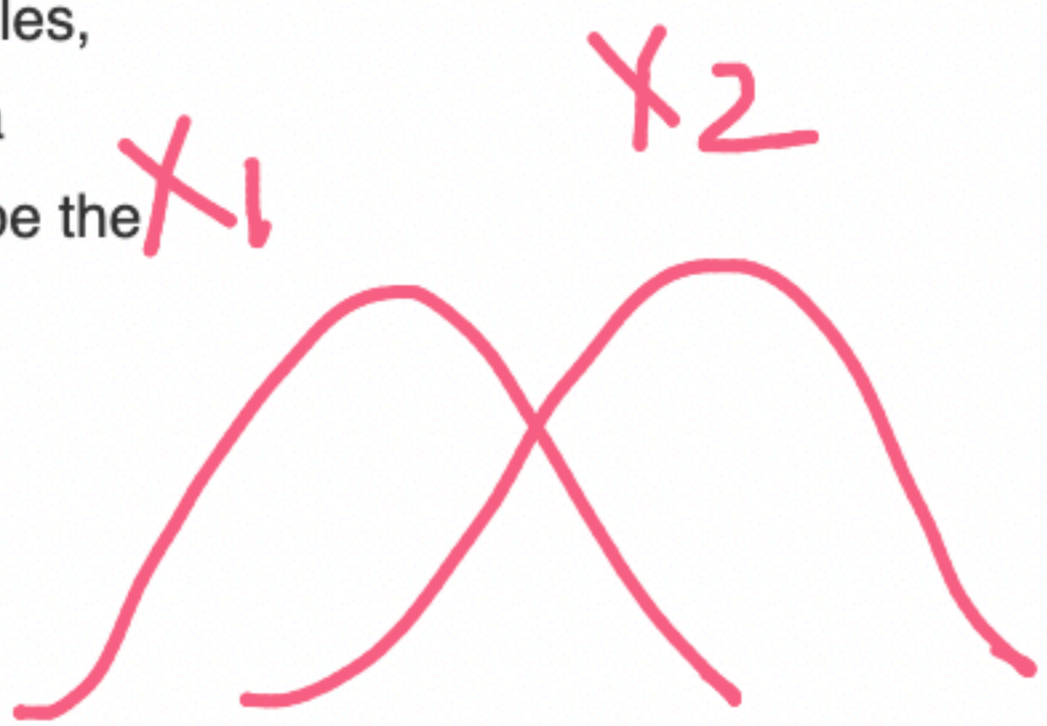
mode : *{'auto', 'exact', 'approx', 'asymp'}, optional*

Defines the distribution used for calculating the p-value. The following options are available (default is 'auto'):

- 'auto' : selects one of the other options.
- 'exact' : uses the exact distribution of test statistic.
- 'approx' : approximates the two-sided probability with twice the one-



If array_like, it should be a 1-D array of observations of random variables, and the two-sample test is performed (and rvs must be array_like). If a callable, that callable is used to calculate the cdf. If a string, it should be the name of a distribution in **`scipy.stats`**, which will be used as the cdf function.



args : ***tuple, sequence, optional***

Distribution parameters, used if *rvs* or *cdf* are strings or callables.

N : ***int, optional***

Sample size if *rvs* is string or callable. Default is 20.

alternative : ***{'two-sided', 'less', 'greater'}, optional***

Defines the null and alternative hypotheses. Default is 'two-sided'. Please see explanations in the Notes below.

$X_1 > X_2$

$X_2 > X_1$

mode : ***{'auto', 'exact', 'approx', 'asymp'}, optional***

Defines the distribution used for calculating the p-value. The following options are available (default is 'auto'):

- 'auto' : selects one of the other options.
- 'exact' : uses the exact distribution of test statistic.
- 'approx' : approximates the two-sided probability with twice the one-

$\left\{ \begin{array}{l} X_2 > X_1 \\ \text{or } X_1 > X_2 \end{array} \right.$

$X_1 \neq X_2$

stic

90 / 90

If array_like, it should be a 1-D array of observations of random variables, and the two-sample test is performed (and rvs must be array_like). If a callable, that callable is used to calculate the cdf. If a string, it should be the name of a distribution in `scipy.stats`, which will be used as the cdf function.

args : *tuple, sequence, optional*

Distribution parameters, used if *rvs* or *cdf* are strings or callables.

N : *int, optional*

Sample size if *rvs* is string or callable. Default is 20.

alternative : *{'two-sided', 'less', 'greater'}, optional*

Defines the null and alternative hypotheses. Default is 'two-sided'. Please see explanations in the Notes below.

mode : *{'auto', 'exact', 'approx', 'asymp'}, optional*

Defines the distribution used for calculating the p-value. The following options are available (default is 'auto'):

- 'auto' : selects one of the other options.
- 'exact' : uses the exact distribution of test statistic.
- 'approx' : approximates the two-sided probability with twice the one-