Assumptions of chi-square test
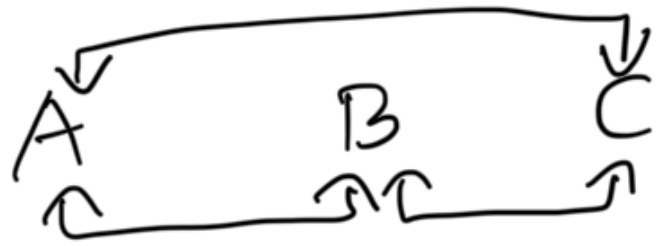
① random sampling

② mutually exclusive groups. ✓✓ ✳

③ <u>observations</u> are independent
samples

<u>ANOVA</u> : Analysis of Variance.

product vs income

idea : apply multiple t-test

A   B   C   3 - test.

10 categories : $^{10}C_2$ = 45 t-test

ANOVA :

H_0 : the means of all the group are same

H_a : at least one group has a diff mean.
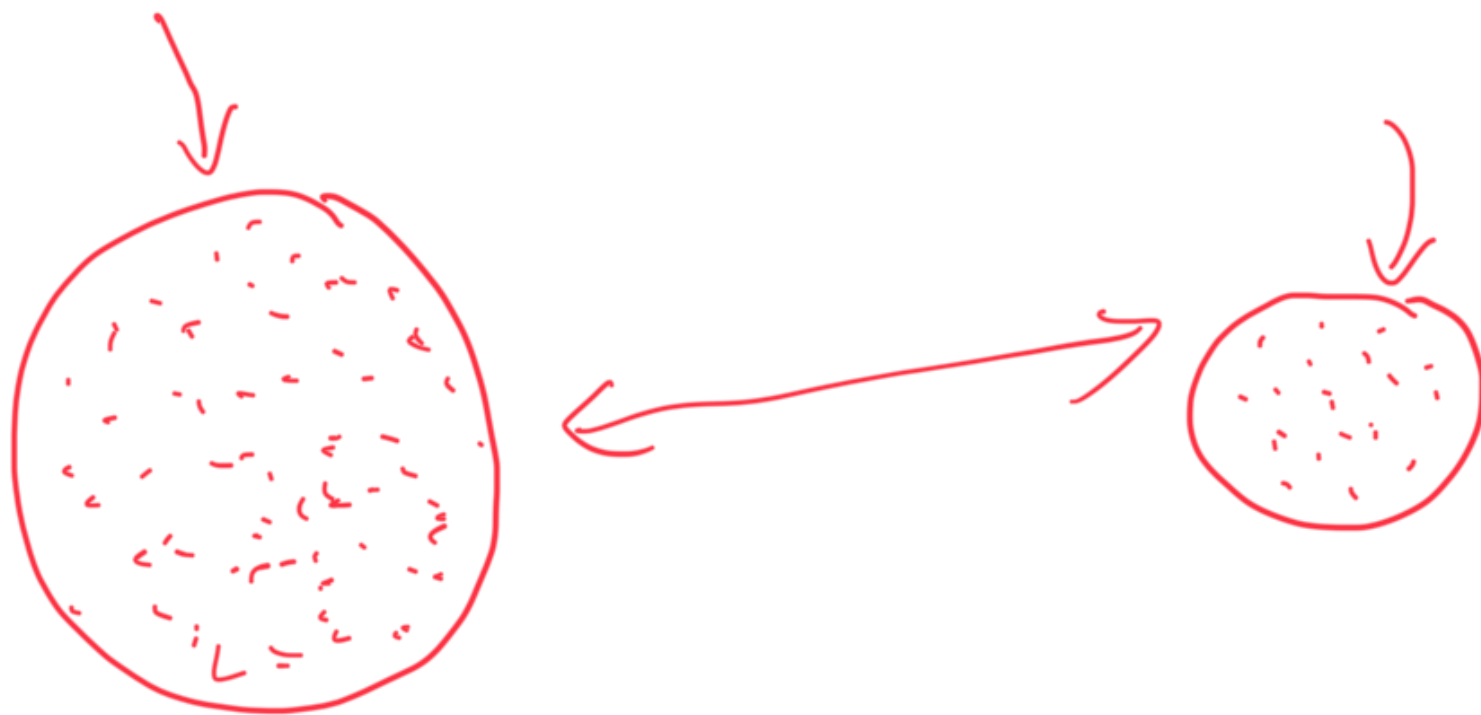
↓

Product :

α

①

②
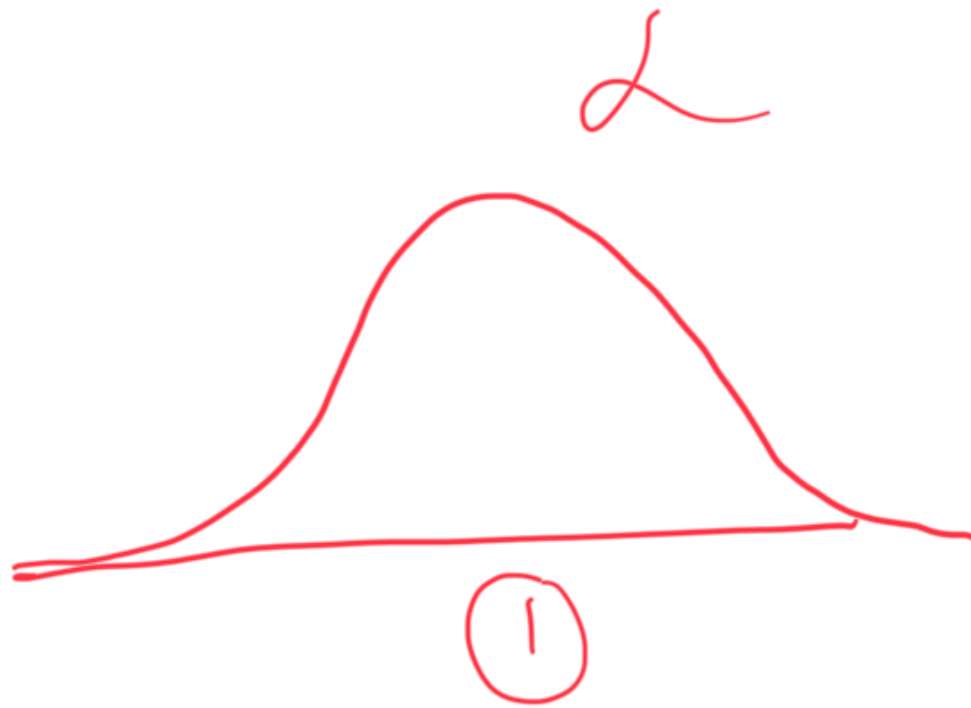
ANOVA

① variance <u>within</u> groups ↓

② variance <u>between</u> groups. ↑

F ratio : if groups are actually different

$$= \frac{\text{variance between groups}}{\text{variance within groups.}}$$

Assumptions of ANOVA

→ data is gaussian ✓   tested: shapiro-wilk

→ samples are independent ✓

→ variance of all the groups is same   levenc-test

Kruskal – Wallis test   (no normality assumption)

H_0 : no diff in the ~~mean~~ median of groups.

H_a : at least 1 is different from others.

→ robust to outliers.
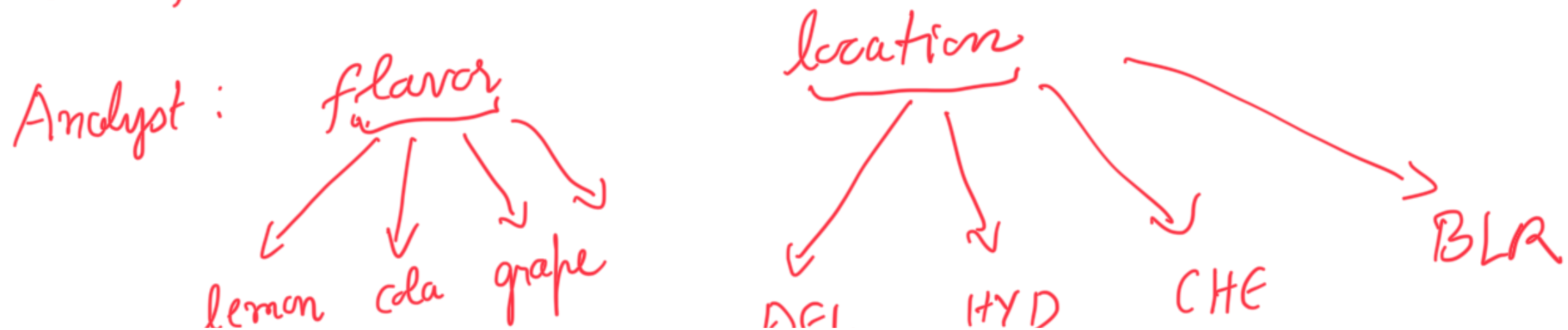
Levene's test [variance]

p-value

H_0 : the variance across diff groups is same

H_a : variance is not the same.

---

independ var : gender

Q: factors that influence sales of a drink

Analyst : flavor   location

flavor → lemon cola grape

location → DEL HYD CHE BLR

EXP : multiple 1-way ANOVA

2-way ANOVA : simultaneously assess
multiple → ind variables.
location ↗ flavor

dependent var
↓
sales.

Main effects

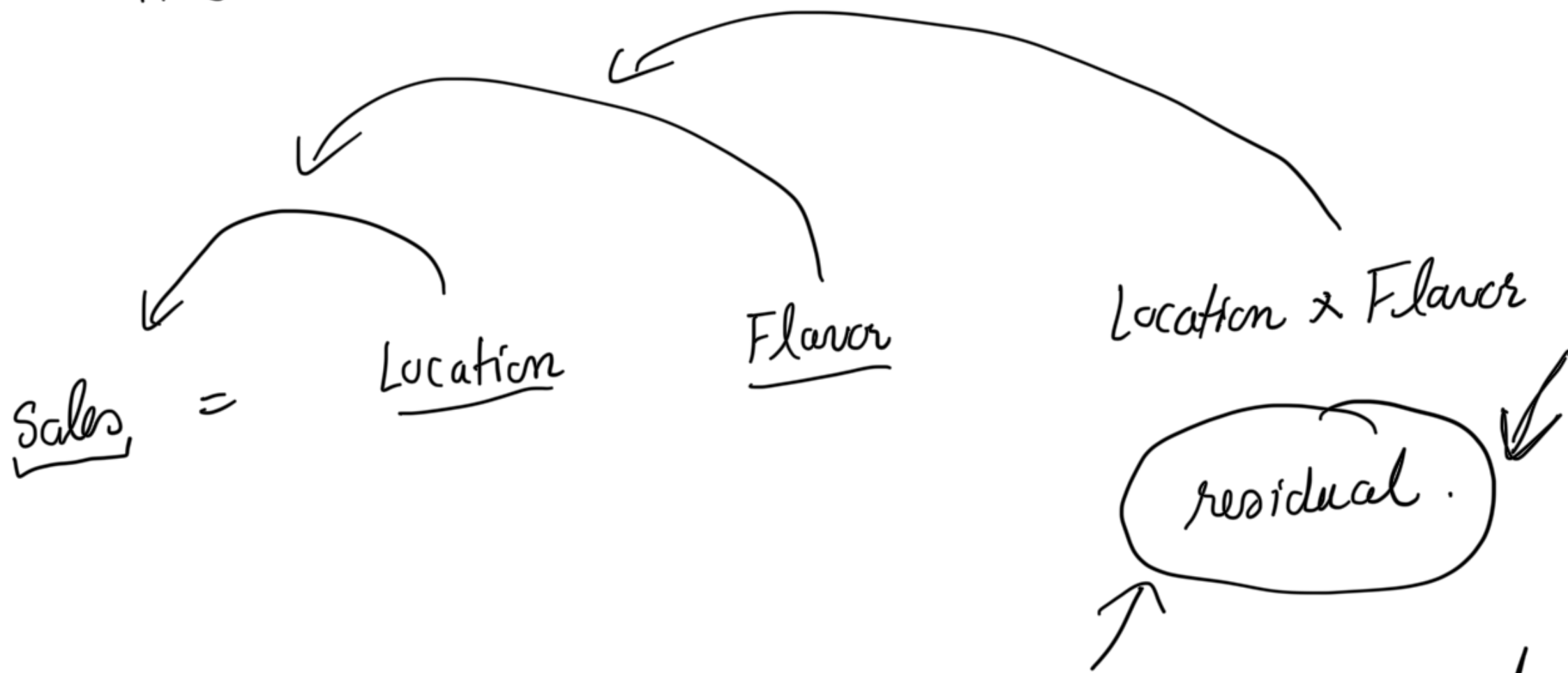H_0 : 1: no effect of flavors on sales.
2. " " " location on sales.

flavour has different effect

Ha: 1. at least 1 flavor "   "   "

2. "   "   " location "   "   "

interaction effect

Ho : no effect because of interaction on sales.

Ha : at least 1 interaction has an effect on sales

$$Sales = Location + Flavor + Location \times Flavor + \text{residual}$$

variation that cannot be explained
by the independent variables

KS- test

Sit: M_1, M_2

recovery time

can we apply
t-test or z-test ??

assumption normality.

KS- test   no assumptions required

compares the distribution
(not means)

•   •
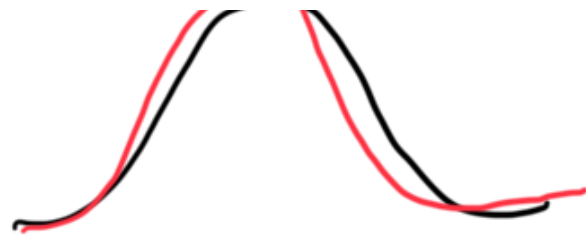
H_0: dist of X and Y is same

H_a: dist of X and Y are diff

$$T_{KS} = |CDF(X) - CDF(Y)|$$

KS test is right tailed

H_0 : dist are same

H_a : $X > y$

A/B testing

10000 → randomly

→ 50% ⟹ 1 ad revenue

↳ 50% ← 2 ads ↑ watch
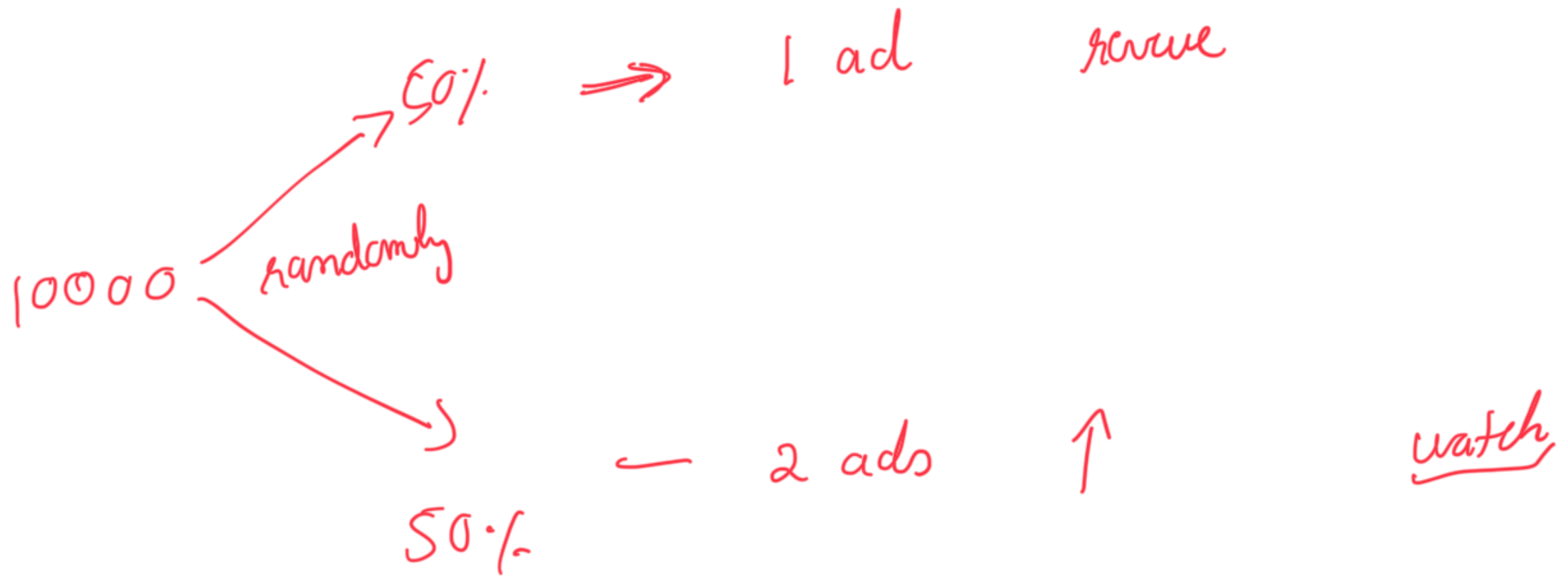
revenue ↑ //

lost watch time

1. <u>Treatment Group</u>,

2. <u>Control group</u>.