Two-Way Anova

Questions You are a DS at Coca-Cola India
You want to check the Factors
influencing sale of Drinks
You have decided on two Factors

1) Flavor (Lemon, Cola, Orange ...)

2) The Location (North, south, east, west)

One-way Anova independently can handle each of the above Factor
However if we want to check for interaction Effect Between the two Variable we need two-way ANOVA

Two-way Anova

1) Independent affect of Var (Main effect)
2) Interaction Effect as well

Hypothesis in 2-way

Main Effects H_0

1. No significant impact of 'flavor' on Sales
2. No significant impact of 'Region' on Sales

Main Effects H_a

1. significant impact of 'flavor' on Sales
2. significant impact of 'Region' on Sales

Interaction Effects H_0

1. No interaction effect of 'Flavor and Region' on sales (Impact of 'Flavor on sale doesnot depend on Region')

Interaction Effects H_a

1. interaction effect of 'Flavor and Region' on sales (Impact of 'Flavor on sale depends on Region')

KS Test

Questions

Imagine you're working as a data analyst for a pharmaceutical company, and your company is researching the effectiveness of two different medicines, Medicine M1 and Medicine M2.

- You want to determine whether these two medicines have similar recovery time distributions when administered to patients.
- So, you decide to examine the distribution of recovery times for both medicines to see if there are any significant differences.

$M_1 \rightarrow [x_1, x_2, x_3 \; ---- \; x_{100}]$ $\quad n = 100$

$M_2 \rightarrow [x_1', x_2', x_3' \; ---- \; x_{150}']$ $\quad m = 150$

Can we use Z-test / T-test here?

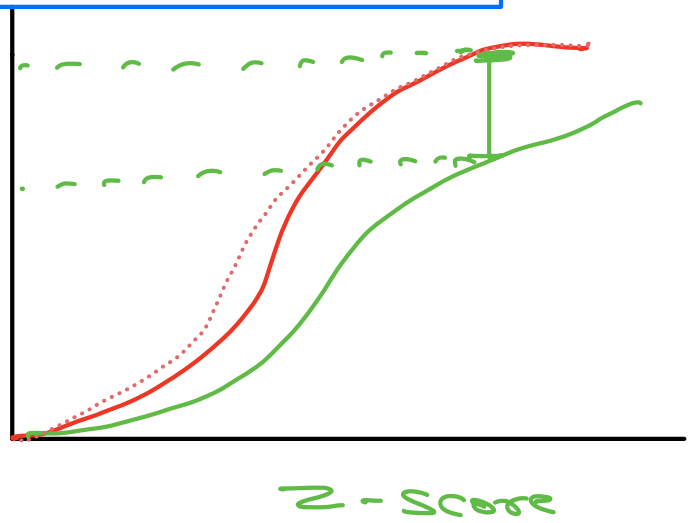
↳ If the task was to check mean recovery time, then Yes

$X : [x_1, x_2 \; ---- \; x_n]$

$Y : [y_1, y_2 \; ---- \; y_m]$

2-Sample Distribution Free Test (Non-parametric)

$$T_{KS} = \sup \left| cdf(x) - cdf(y) \right|$$



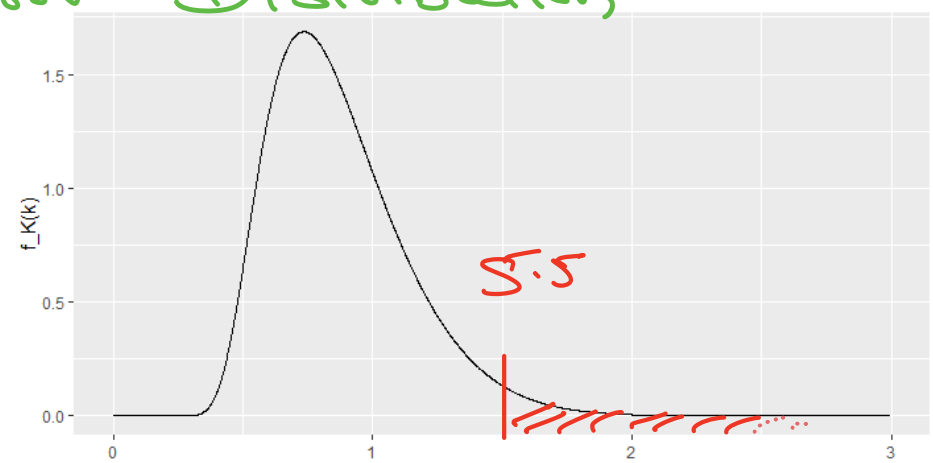
H_0 : X and Y follow Same Distribution

H_a : X and Y follow different Distribution

Test - statistic (K-S) ⟶ P-value (Compare with α)

↓

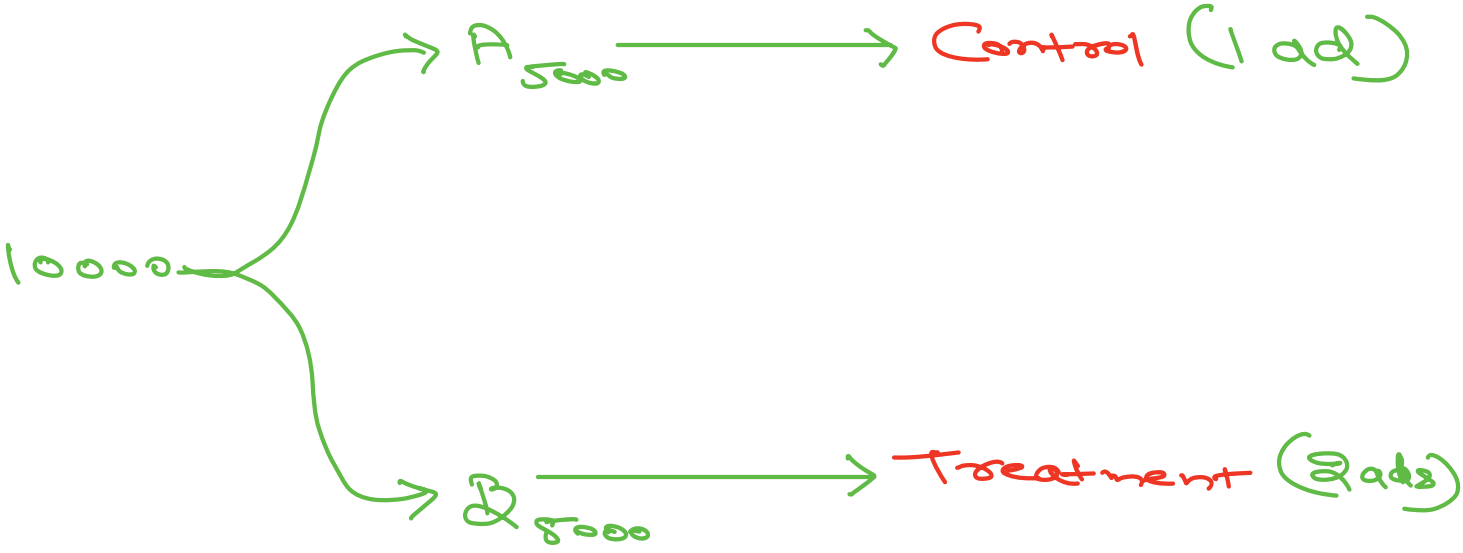
Kolmogorov Distribution



Right Tailed Test

(0.95) CL ⟶ k value (Critical)

AB Testing



H_0 : Mean Watch Time A
=
Mean Watch Time B

H_a : Mean Watch Time A
≠
Mean Watch Time B

Parametric Test vs Non Parametric Test

Parametric Test: Will have some assumption on Distribution of underlying population

Non-parametric Test: Will have few or No assumption on Distribution of underlying population

A	B	C
Test	**Type**	**Reason**
One Sample Z-Test	Parametric	Known population standard deviation, normally distributed data
Two Sample Z-Test	Parametric	Known population standard deviations, normally distributed data
One Sample T-Test	Parametric	Normally distributed data
Two Sample Independent T-Test	Parametric	Normally distributed data, equal variances
Paired T-Test (Dependent)	Parametric	Normally distributed population differences
One Sample Z-Test Proportion	Parametric	Known population proportion, large sample size
Two Sample Z-Test Proportion	Parametric	Known population proportions, large sample size
Chi-Square Test	Non-parametric	Tests association between categorical variables
One Way ANOVA	Parametric	Normally distributed data, equal variances
Kruskal-Wallis Test	Non-parametric	Alternative to ANOVA when assumptions are not met
Shapiro-Wilk Test	Parametric	Checks normality of data
Levene's Test	Parametric	Checks homogeneity of variances
Two-Way ANOVA	Parametric	Extends one-way ANOVA to study two factors
KS-Test (Kolmogorov-Smirnov Test)	Non-parametric	Compares distributions of two samples
A/B Testing	Parametric or non-parametric	Depends on the specific metric and data