Hypothesis Testing

Advance

Recap

1. Z-Test
   - One Sample Z-Test — compare sample mean ($\bar{x}$) with population mean (μ)
   - Independent Sample Z-Test — compare avg across two ind. groups

   [population std. deviation is known and sample size $n > 30$]

2. Z-proportion Test
   - one proportion Z-Test — compare sample proportion with population prop.
   - Two proportion Z-Test — compare proportion across two indep. groups

3. T-Test (pop. std. dev is unknown and n (sample size) < 30)
   - one sample T-Test → compare sample mean with pop. mean
   - Ind. sample T-Test → compare avg. across two indp. groups
   - paired sample T-Test → compare avg of a group in two different time segment (Before - After)

4. chi-Square
   - goodness of fit → If a series is uniformly distributed or not?
   - Test for independence → Test if two categorical variables are dependent or not?

5. Anova
   - one way → to compare avg across more than two groups (avg height) — mothers, fathers, childers
   - two way

Hypothesis Testing

1. formulate the Hypo. — NULL H0, Alternative H1/Ha
2. Choose the distribution
3. You compute p-value
4. compare p-value with significance level alpha
5. Conclusion.

test if the attendance was uniform across all locations

Loc.	Attendance
Delhi	
BAG	
CHN	
Indore	
Lucknow	

1. A company produces chip packets and claims that each packet comes with 100g of chips with a standard deviation of 5g. The Food Department doubts the claim of the company after receiving multiple complaints from customers. To test this, they have collected 50 chip packets and found that the average weight of chips was 95g.

Z-Test population mean $\mu = 100g$ $\sigma = 5g$ $n = 50$ $\bar{x} = 95g$

one sample Z-Test

2. The school principal of Dholkpur believes that 80% of students in his school are excellent with Vedic math. Bheem doubts the principal's belief and took 40 students from different classes, giving them 10 Vedic math problems. Out of the 40 students, 35 were able to pass the test with good marks. Is the principal correct about his belief?

one sample Z-prop. Test $p = 0.8$ (population prop.) $\hat{p} = \frac{35}{40}$ (sample propor)

3. The average weight of female and male students in a class of 50 was found to be 50kg and 55kg, respectively. Can we test if the weight differs between female and male students? (There were 25 male and 25 female students in the class.)

Independent sample T-Test avg(weight) — males, females

4. Dr. Jackal is working on a new formula that can make Shaktimaan lose his memory. In order to test his formula, he asked his fellow criminal friends to participate in this noble cause. Ten criminals participated, and their IQ levels before taking the medicine and after taking the medicine were recorded. What kind of test should Dr. Jackal use?

paired sample T-Test [comparing avg before and after the test]

5. Noddy believes that the number of weather conditions depends on the color of hat he wears. There are four weather conditions, and he has five different colored hats. Which test should Noddy be using?

Chi-Square test of Independence.

Note: Refer to the below table to answer all the questions.

Anova
one way
Two way
KS-Test
AB-Test

1. Refer to the table and identify the variable for one way Anova and the two way Anova.

one categorical column
[one continuous column]
avg

avg(Sales) — orange, Lemon, Cola → flavour → Sales → avg
avg(Sales) — East, west, north, South → location → Sales → avg
one-way Anova

one way Anova →
(1) Test if the flavour of the product is impacting sales or not?
(2) Test if the location of the product is impacting sales or not?
} one-way Anova
(3) Test if location and flavour is impacting sales or not?

avg(s) ⇌ avg(s) ⇌ avg(s) } If sales is dependent on flavour
orange, Lemon, Cola

avg(s) ⇌ avg(s) ⇌ avg(s) ⇌ avg(s) } If sales is dependent on location
East, west, north, South

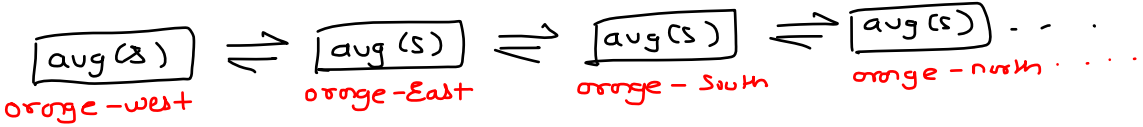
	Flavour	Location	Sales
0	Orange	West	141
1	Lemon	West	178
2	Orange	West	170
3	Orange	East	76
4	Lemon	East	170
5	Lemon	East	165
6	Orange	North	52
7	Cola	West	152
8	Orange	East	186
9	Orange	East	111
10	Orange	North	100
11	Orange	East	108
12	Lemon	North	167
13	Orange	South	145
14	Cola	East	162

Note: Refer to the below table to answer all the questions.

Two categorical columns [one continuous col.] → we compute avg

	Flavour	Location	Sales
0	Orange	West	141
1	Lemon	West	178
2	Orange	West	170
3	Orange	East	76
4	Lemon	East	170
5	Lemon	East	165
6	Orange	North	52
7	Cola	West	152
8	Orange	East	186
9	Orange	East	111
10	Orange	North	100
11	Orange	East	108
12	Lemon	North	167
13	Orange	South	145
14	Cola	East	162

1. Refer to the table and identify the variable for one way Anova and the two way Anova.



The comparison is based on the interaction of two categorical variables

Test if sales is dependent on flavour & location or not?

Note: Refer to the below table to answer all the questions.

1. Refer to the table and identify the variable for one way Anova and the two way Anova.

Main Effect Interaction Effect

Main Effect [one to one relationship] one categorical (flavour or location)
one continuous (sales)

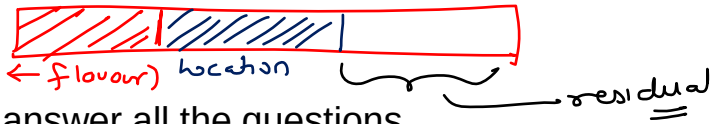
flavour H_0 → the avg sales across all flavour is the same.
H_a → the avg sales across all flavour is not same / or different.

Location

Interaction Effect [two to one relationship] two categorical (flavour and location)
one continuous (sales)

H_0: the avg sales is not impacted by flavour and location
H_a: the avg sales is impacted by flavour and location

	Flavour	Location	Sales
0	Orange	West	141
1	Lemon	West	178
2	Orange	West	170
3	Orange	East	76
4	Lemon	East	170
5	Lemon	East	165
6	Orange	North	52
7	Cola	West	152
8	Orange	East	186
9	Orange	East	111
10	Orange	North	100
11	Orange	East	108
12	Lemon	North	167
13	Orange	South	145
14	Cola	East	162



Note: Refer to the below table to answer all the questions.

1. Refer to the table and identify the variable for one way Anova and the two way Anova.

	X	$(x-\bar{x})^2$
a	5	$(5-7.4)^2 = 5.76$
b	7	$(7-7.4)^2 = 0.16$
c	10	$(10-7.4)^2 = 6.76$
d	12	$(12-7.4)^2 = 21.16$
e	3	$(3-7.4)^2 = 19.36$

avg(x) = 7.4

Σ 53.2 → sum of sq.

$\Sigma \frac{53.2}{5}$ = var

	Flavour	Location	Sales
0	Orange	West	141
1	Lemon	West	178
2	Orange	West	170
3	Orange	East	76
4	Lemon	East	170
5	Lemon	East	165
6	Orange	North	52
7	Cola	West	152
8	Orange	East	186
9	Orange	East	111
10	Orange	North	100
11	Orange	East	108
12	Lemon	North	167
13	Orange	South	145
14	Cola	East	162

one way Anova

1. If the cholesterol is influenced with Sex of the patient or not?
2. If the cholesterol is influenced by BP-stakes of the patient or not?

Two way Anova

If the cholestrol is impacted by Sex & BP-stakes of a patient or not?

=

$$\text{f-stats} = \frac{\text{variance between groups}}{\text{variance within groups}}$$

variance = OLS sum of sq $\begin{matrix} f \\ L \\ f-L \end{matrix}$

Obs	Cholesterol	Sex	BP_Status
1	194	Male	Normal
2	200	Female	High
3	233	Male	High
4	192	Female	Optimal
5	209	Female	Normal
6	200	Female	High
7	184	Female	Normal
8	228	Female	High
9	150	Female	Normal
10	221	Male	Normal

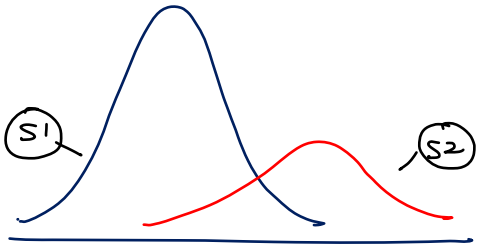
Agenda:

- **Recap of One-Way ANOVA**
- **Two-Way ANOVA**
- **KS - Test**
- **A/B Testing**
- **Parametric vs Non-parametric.**

K-S Test (Kolmogrov – Smirnov Test)

Non-parametric test

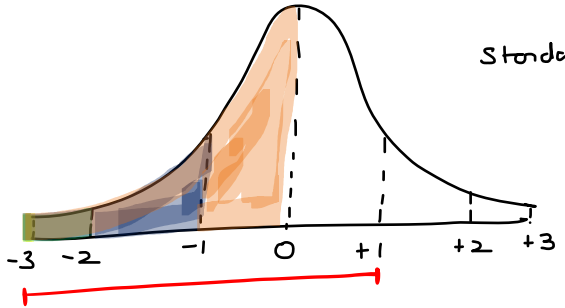
To compare if two series comes from some distribution or not?



is Series ① and ② follow some distribution?

NO

CDF – Cumulative Dist. function

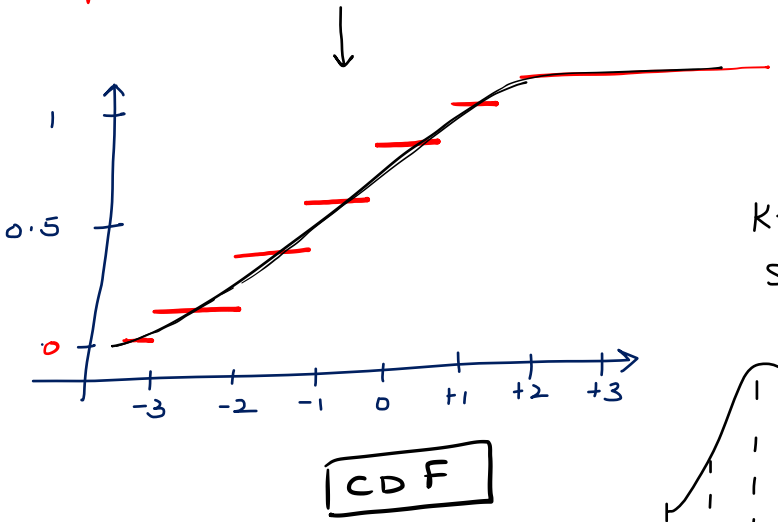


Standard normal Dist

$\mathcal{N}(0, 1)$

H_0: Both are same

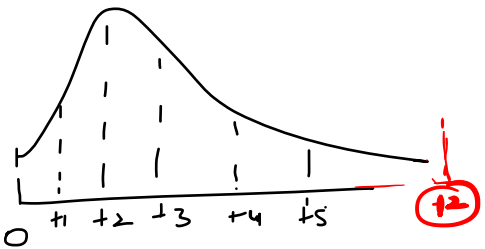
H_a: Different



CDF

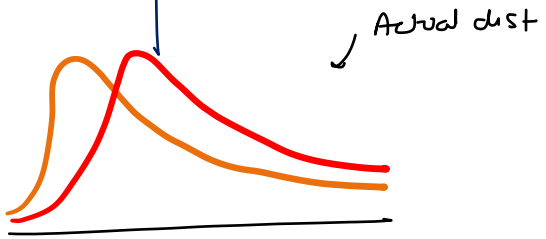
K-S-statics =

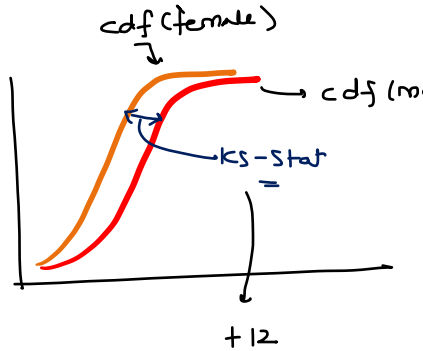
Sup | CDF(x) − CDF(y) |



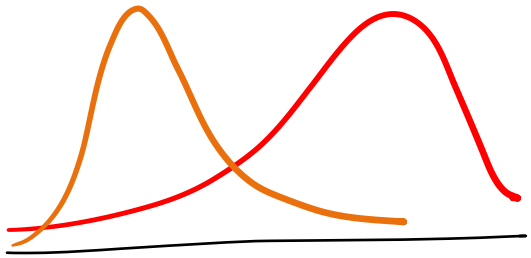
avg

avg





p-value = 0.0000



T-Test
Z-Test
KS-Test

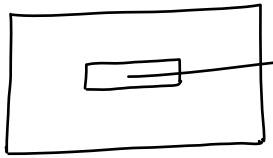
Z-Test
T-Test
→ pop. std.
n-sample
Sample mean
popu Sm...

x	y

KS test

AB-Testing

amazon.in



amazon wants to test if the Buy now button should be of Red or Yellow colour?

exposed to the change

Treatment group (1M customers) → Red colour button

→ Control group (9M customers) → Yellow colour (default colour already present)

Not exposed to the change

avg(Sales) Treat ⇒ avg(Sales) Cont