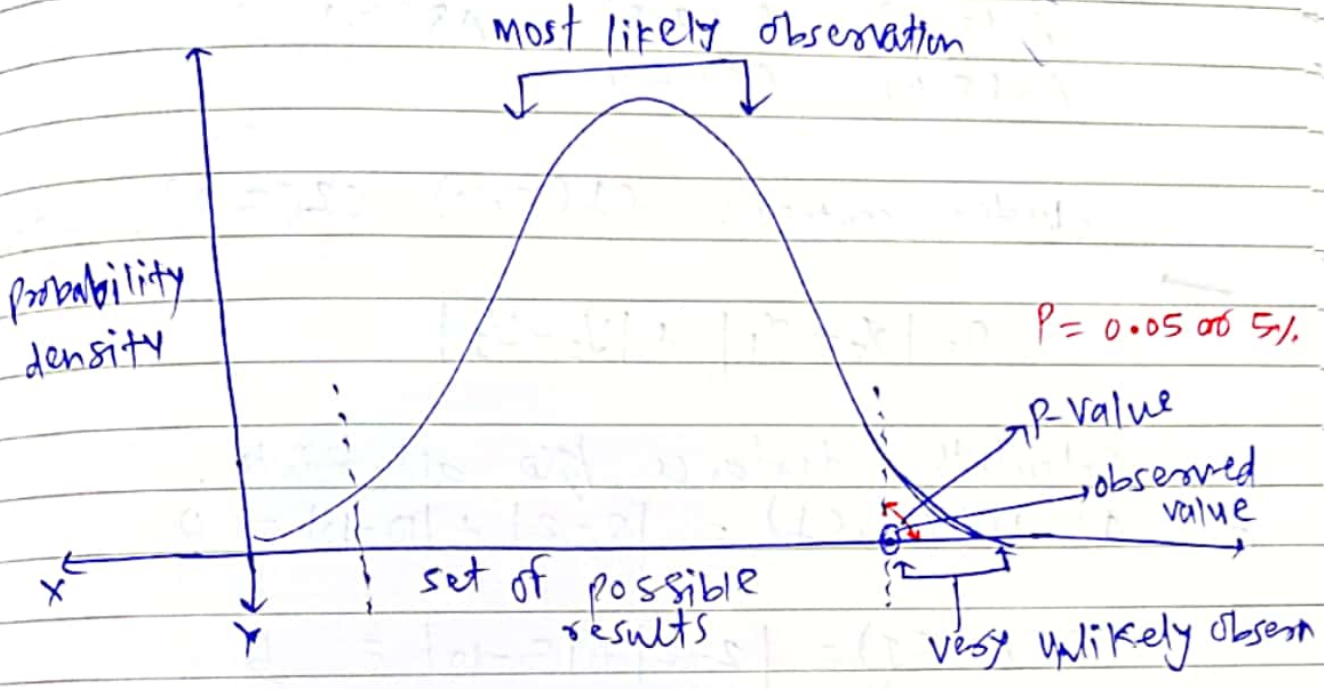
Range of p-value ⇒ $0 \leq \text{p-value} \leq 1$

DOMS | Page No. | Date / /

Significance of p-value :-



* NULL Hypothesis
* alternate Hypothesis

1. Problem statement
2. Data collection & preparation
3. Training & testing split
4. validation state
5. Result

~~P = 0.05 or 5%~~

If $p \leq 0.05$, reject NULL Hypothesis & accept alternate hypothesis

If $p > 0.05$, reject alternate hypothesis & accept NULL Hypo.

NULL Hypothesis Statement : Assumption regarding data.

set of possible results indicate probability density.

Alternate hypothesis is alternative assumption for data, for this diagram we can state alternate hypo. as there is no relation b/w set of possible results & probability density.

based on similar features we categorize them into groups.

DOMS | Page No. | Date / /

K-means clustering:-

Q. Cluster following 8 pt. into 3 clusters in the form of ~~x & y~~ (x, y)

A1(2,10) A4(5,8) A7(1,2)
A2(2,5) A5(7,5) A8(4,9)
A3(8,4) A6(6,4)

cluster centres: C1(2,10), C2(5,8), C3(1,2)

→

$$D = |x_2 - x_1| + |y_2 - y_1|$$

Calculate distance b/w all points.

Ans: 1) D(A1, C1) = |2-2| + |10-10| = 0

2) D(A2, C1) = |2-2| + |5-10| = 5

3) D(A3, C1) = |8-2| + |4-10| = 12
4) D(A4, C1) = 3+2 = 5
5) D(A5, C1) = 5+5 = 10
6) D(A6, C1) = 4+6 = 10
7) D(A7, C1) = 1+8 = 9
8) D(A8, C1) = 2+1 = 3

D(A1, C2) = 3+2 = 5	D(A1, C3) = 9
D(A2, C2) = 3+3 = 6	D(A2, C3) = 4
D(A3, C2) = 3+4 = 7	D(A3, C3) = 9
D(A4, C2) = 0+0 = 0	D(A4, C3) = 10
D(A5, C2) = 2+3 = 5	D(A5, C3) = 9
D(A6, C2) = 1+4 = 5	D(A6, C3) = 7
D(A7, C2) = 4+6 = 10	D(A7, C3) = 0
D(A8, C2) = 1+1 = 2	D(A8, C3) = 10

Based on minimum distance choose pt. belong to cluster

Jan 15 | Page No.
Date / /

Formation of Table :

Given pts	Dist from C1(2,10)	from C2(5,8)	from C3(1,2)	pt. belong to cluster
A1	0	5	9	C1
A2	5	6	4	C3
A3	12	7	9	C2
A4	5	0	10	C2
A5	10	5	9	C2
A6	10	5	7	C2
A7	9	10	0	C3
A8	3	2	10	C2

List cluster centres:-

C1(A1(2,10))

C2(A3(8,4), A4(5,8), A5(7,5), A6(6,4), A8(4,9))

C3(A2(2,5), A7(1,2))