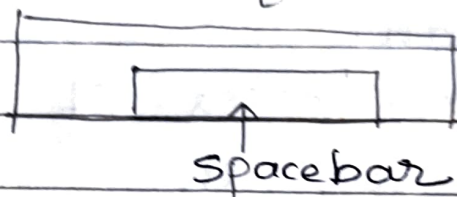


## P-Value

Date: 25/2/20 Page No.:

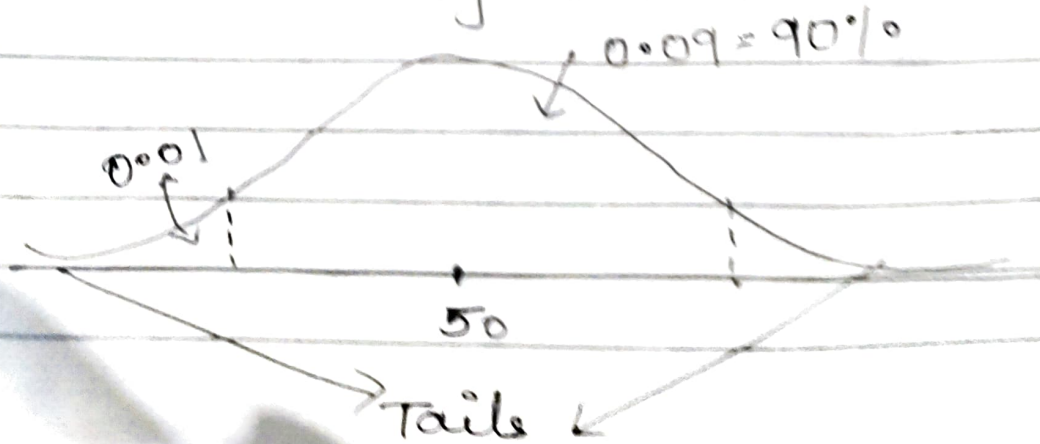
P-value (Significance Value) specifies the probability of null hypothesis being true. Null Hypothesis ( $H_0$ )  $\rightarrow$  Treats everything as same or equal.

For eg:



We use spacebar frequently in the keyboard. Most of the times we hit spacebar in the middle region of the key. The chances of hitting space bar at its extreme corners is very less. Hence we can say that if we perform an experiment of hitting space bar 100 times, the chances of hitting middle region is 90% and chances of hitting extremes are 10%. Here p value will be 0.09 and 0.01 respectively.

If we try to plot this experiment we get the curve something like this:

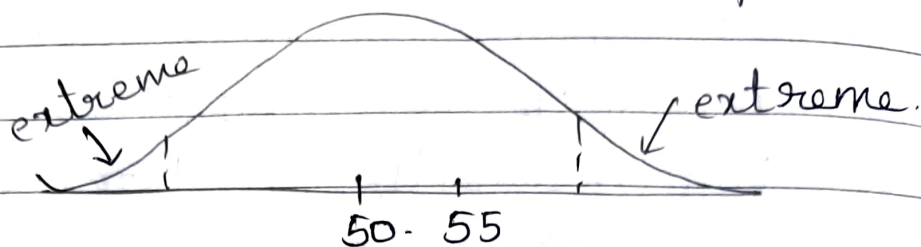


$H_0$ : The coin is fair  $H_1$ : The coin is unfair  
For every 100 tosses, if I get 50 time head it is fair coin.

Date:

Page No.:

In this case I can ideally represent it as

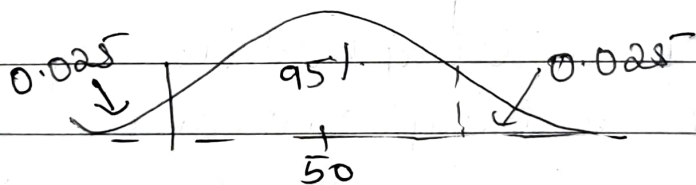


50 being the mean.

For first 100 exp tosses if I get 55 times head it falls in the region and is nearer to mean 50. Hence we accept the Null hypothesis. p value here falls within the significance region

p-value is defined or given by domain expert

If p-value is  $0.05 \rightarrow 0.025$  each on extreme



ends and 95% will be in the middle region

If second 100 tosses of a coin ends up having 20 heads, it falls in the left extreme tail. Here  $p \text{ value} < 0.025$ . Hence null hypothesis is rejected.

## Types of Test

consider the following table:

Date:

Page No.:

Gender	Age group	Height (cm)	Weight (kg)
M	20 Adult	1.4	60
F	19 Adult	1.2	55
M	Elderly	1.4	72
M	child	1.0	25
M	Adult	1.3	58
F	child	1.2	18
M	Elderly	1.25	82
F	Elderly	1.3	67

H<sub>0</sub>: If there is any Variation in the proportion of male & female based on gender

H<sub>0</sub>: There is no Variation

H<sub>1</sub>: There is a Variation

For one categorical feature, we apply one sample proportion test.

Here  $p \leq 0.05$  (Reject H<sub>0</sub>)

If there is any Variation in the proportion of male & female based on Age group. Here there are two categorical variables.

We apply Chi-square test.

If there is any Variation in mean height based on previous sample? Here we apply T-test for one numerical feature/variable.



If we consider two continuous features (numerical variables) we apply co-relation and T-test (pearson lies within  $-1$  to  $1$ ) Date:

Page No.:

For one numerical & one-category - T-test (two groups) ~~ANNOVA~~

For one numerical & one-category (diff groups) ANNOVA