DAY-9 DATA SCIENCE
STATISTICS

P-value

⇒ It is the probability for the Null Hypothesis to be true

* A P-value less than 0.05 (typically ≤ 0.05) is statistically. It indicates strong evidence against the null Hypothesis, as there is less than 5% Probability that the null value is correct (and the Results are random). Therefore we Reject the Null Hypothesis & Accept the Alternative Hypothesis.

Chi-Square test

* Chi-Square test claims about Population Proportion.

* It is non Parameter test which are Performed upon categorical Data

1. NOMINAL 2. ODINAL

$$\chi^2 = \sum \left(\frac{F_o - F_e}{F_e} \right)^2$$

F_o = Observation value

F_e = ~~Actual~~ expected value

Anova → Analysis of variance | F-test

* For comparision of more than two population or population having more than two sub groups. we use Anova technique

* Statistical Technique developed to study significance of difference of mean of 2 or more than 2 samples.

$$F_{(N,D)} = \frac{\text{variation b/w sample}}{\text{variation within sample}}$$

Degree of freedom 1 for denominator & 1 for Numerator

Technique of Anova

1. One way Anova
2. Two way Anova

Bernoulli Distribution

⇒ The outcomes of this Distribution is Binary.

ex:- Yes No

0 1

True, False

Practical Example

Tossing a coin

Heads or tail

$P = P_{(H)} = 0.5$ $Q = P_{(T)} = 0.5$

$P = 1 - q$ $Q = 1 - P$

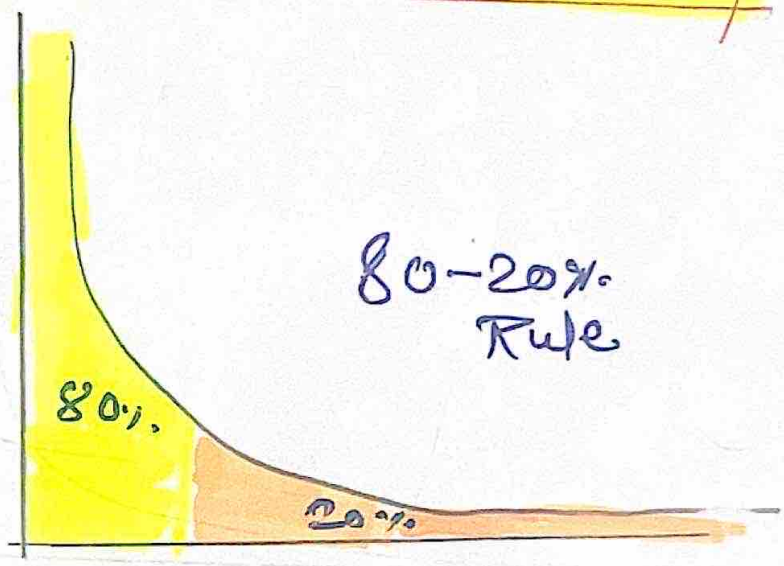
* Most of the classification Problems are Bernoulli Distribution.

Binomial Distribution

In a Binomial Distribution, the Probability of getting a Succes must remain the same for the trails we investigating.

For example, the Probability of tossing a coin is 0.5 H/T for every trail or experiment we conduct.

Power Law Distribution



80-20% Rule

example

* 80% Sales are coming from 20% of over all Product

* 80% of MS windows crash are because 20% of all the buy

A Power Law is a Relationship between two Quantities where if :-
A Relative change in one Quantity results in a Proportional change in other Quantity.

example

⇒ 80% of DataScience we 20% of Software

Pareto Distribution is kind of Powerlaw Distribution.

Q. How DO we convert Pareto Distribution to Normal Distribution

⇒ Box Cox