7 day

① P value & Significance value { 7080 ⇓ KRISHIO → 10% }

② Distribution

③ Central Limit Theorem

④ Bernoulli's Distr Log Normal Distr

⑤ Binomial Distr

⑥ Pareto's (Pareto) Distr {Power Law}

⑦ F Test (ANOVA) → 1 hr → upload a seperate video.

① P value & Significance

↳ Derive the p value (We will see how to)

Eg 1:

Q) The average weight of all residents in Bangalore city is 168 pounds. With a standard deviation 3.9. We take a sample of 36 individuals and the mean is 169.5 pounds. C.I : 95%

We can use Z-test here as n sample is given and std is given

Ans) $\mu = 168$ $\sigma = 3.9$ $\bar{X} = 169.5$ $n = 36$ $\alpha = 0.05$

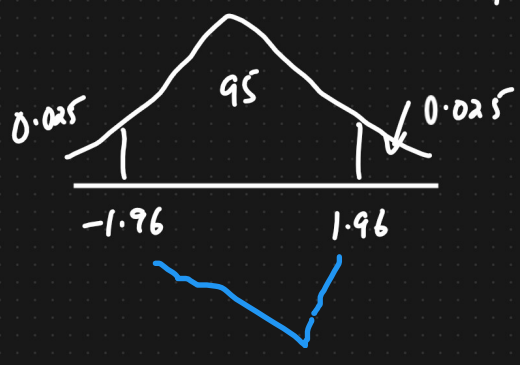
① $H_0 = \mu = 168$ (null hypo)

$H_1 : \mu \neq 168$ (alt hypo)

② $\alpha = 0.05$

③ Decision boundary

$1 - 0.025 = 0.9750$ ⇓



From z-table

Two tailed test as it can be > or < than 168

(5) Z Test Now use z-test formula

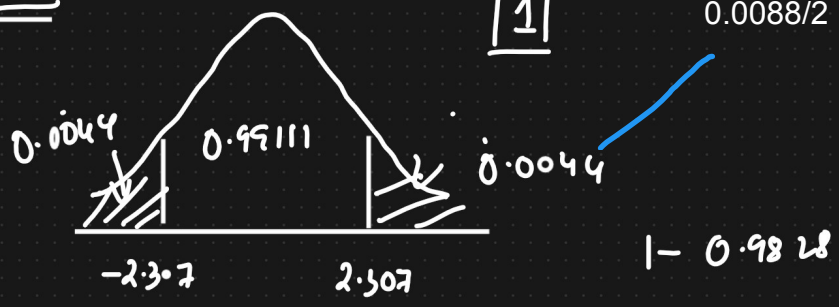
$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{169.5 - 168}{\frac{3.9}{\sqrt{36}}}$$

$$= \frac{1.5}{3.9} \times 6$$

$$= 2.307$$

(6) $Z = 2.307 > 1.96$ Reject the Null Hypothesis

P value



1 − 0.9828

Actually, we get 0.9828 from z-table but we did the calculation using 0.99111 here, but intution is same; we have done calculation in next page wrt 0.9828

Now to get side areas:

$1 - 0.99111 = 0.0088$

Wrt z-score 2.307, we will find area under the curve using z-table we get 0.99111

P value = 0.0044 + 0.0044

= 0.0088

(Now to get P value we will add areas of both ends)

Significance value

Pvalue < 0.05

0.0088 < 0.05 → Reject the Null Hypothesis

Using Pvalue we can also verify that null hypothesis will be rejected,

Here p_value is < significance value, therefore we reject null hypothesis

Note:

②

P value $\leq$ Significance value

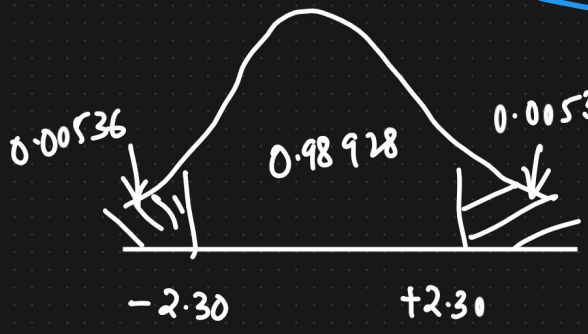
$\Downarrow$

Reject the Null hypothesis

P value $>$ Significan value

$\Downarrow$

Fail to Reject the Null Hypothesis



$Z = 2.30$

$1 - 0.98928 =$

$2.30 > 1.96$ {Reject the Null Hypothesis}

P value $= 0.00536 + 0.00536$

$= \_\_\_\_\_ \leq \alpha \Rightarrow$ Reject the Hypothesis

Eg 2:

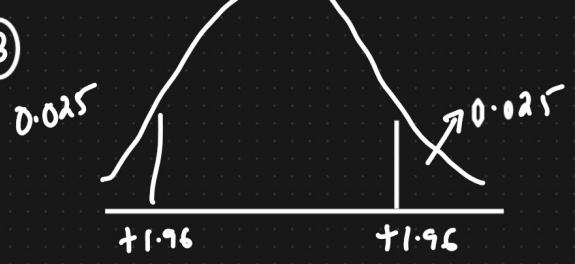
② Average age of a College is 24 years with a standard deviation 1.5. Sample of 36 student students mean is 25 years. With $\alpha = 0.05$ C.I : 95%, do the age vary?

Ans) ① H_0 $\mu = 24$ $\sigma = 1.5$ $n = 36$ $\bar{x} = 25$ $\alpha = 0.05$

H_1 $\mu \neq 24$

② $\alpha = 0.05$

③



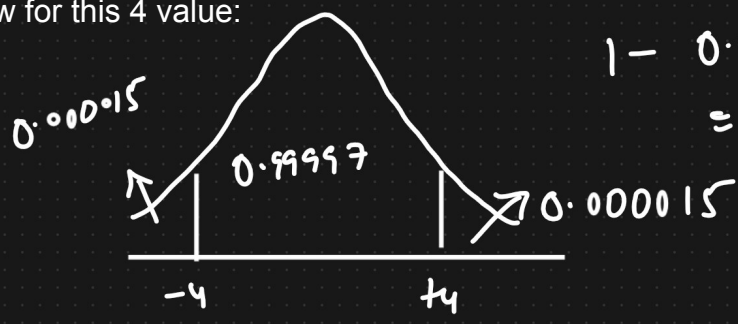
④ $Z\text{-score} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{25 - 24}{1.5} \times 6$

$= \dfrac{1 \times 6}{1.5}$

$= 4$

$4 > 1.96$ Reject Null Hypothesis

Now for this 4 value:



from z-table

$1 - 0.99997$

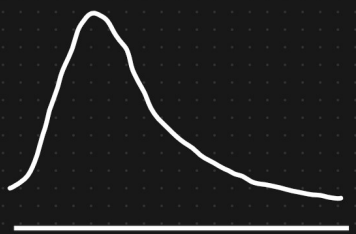
$= 0.00003$

$p\text{value} = 0.000015 + 0.000015$

$= 0.00003$

pvalue $\leq$ Significance {Reject the Null Hypothesis}

② Log Normal Distribution



Eg:- ① Wealth Distribution

② People writing big comments

Here it means that most of the people have less or medium wealth and there are only few which fall under wealthy criteria

{ $y \sim$ Log Normal. Distr }

If a random variable y that belongs to log normal distribution, if we apply log(y), then it should follow normal distribution - if it satisfies this condition, we can say it follows log normal distribution

log(y) ⟶ Normal Distribution

④ Bernoulli's Distribution

In Bernoulli's distribution there are 2 outcomes: 0 or 1, but when we need to focus on probability here, we need to define two values p and q

2 Outcomes | 0 or 1 | Single Trial

Tossing a coin

P = 0.5

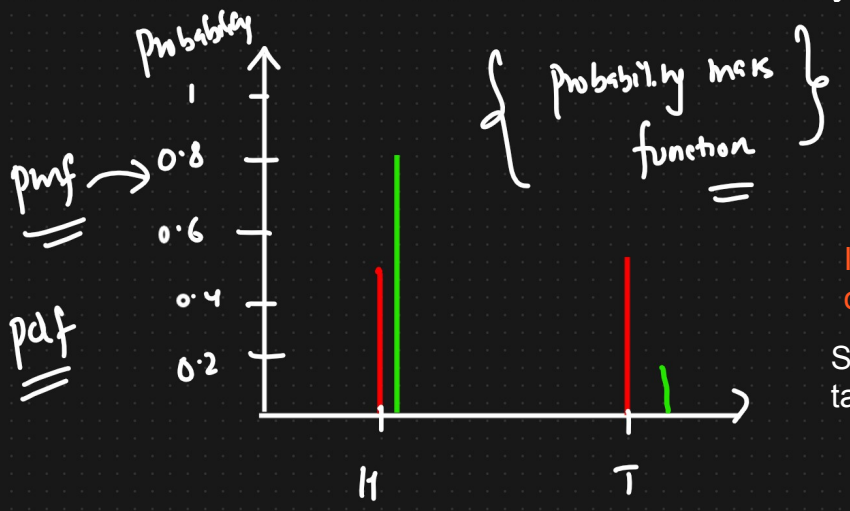
P(H) = 0.5 = P

q = 1 − P = 0.5

Do not have a fair coin

P(H) = 0.3 = P

P(T) = q = 1 − P = 1 − 0.3 = 0.7



Probability density function is for conitnuous variable, here we have categorical variable so this probability mass function

In this graph, when prob of heads is 0.5 then prob of tails is also 0.5 (i.e 1-p)

Suppose, when prob of heads is 0.8 then prob of tails is 0.2

⑤ Binomial distribution

Above we discussed about single trial, when we have multiple trials, it becomes binomial distribution

Every ⟶ Bernoulli distribution

Multiple Trail

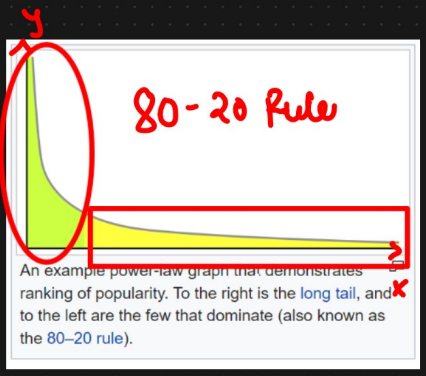
↓

P(H) = 0.5 P(H) = 0.6 - — - - - - .

P(T) = 0.5 P(T) = 0.4

Pareto distribution

- It is not a gaussian distribution
- Example of pareto distribution is power law distribution



An example power-law graph that demonstrates ranking of popularity. To the right is the long tail, and to the left are the few that dominate (also known as the 80–20 rule).

Eg: ① 80% of the wealth is distributed with 20% of the people

② 80% of the company project are done by 20% of the people in a team

③ 80% of sales is done by the 20% of famous product.

④ 80% of the Cricket match is won by 20% of the team.

Power law distribution

Log Normal Distribution

If we extend this graph

we get log normal dist

Lognormal Distribution

→ Transformation

We can lalso watch this log normal distribution to normal distribution, for it watch Krish Naik's data transformation video: https://www.youtube.com/watch?v=3gfhbXt9TcQ

⑤ Central Limit Theorem

If we have some distribution either normal or not, and we take multiple samples such that n>=30.

And for every sample we start calculating their mean

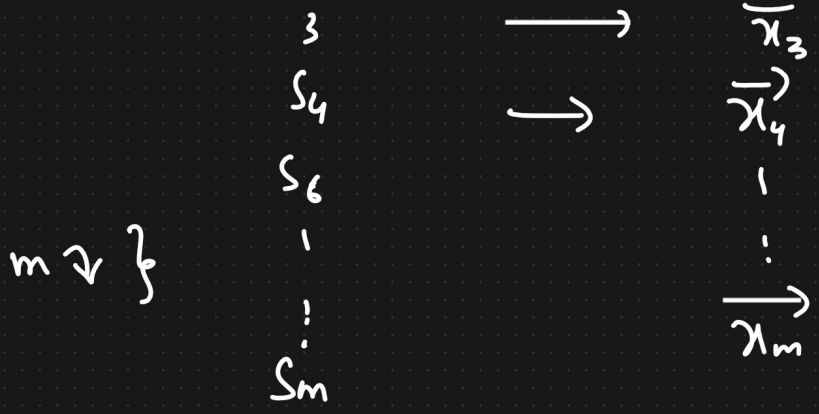
Sample size → $n \geq 30$

(More greater than equal to 30, more central limit theorem holds)

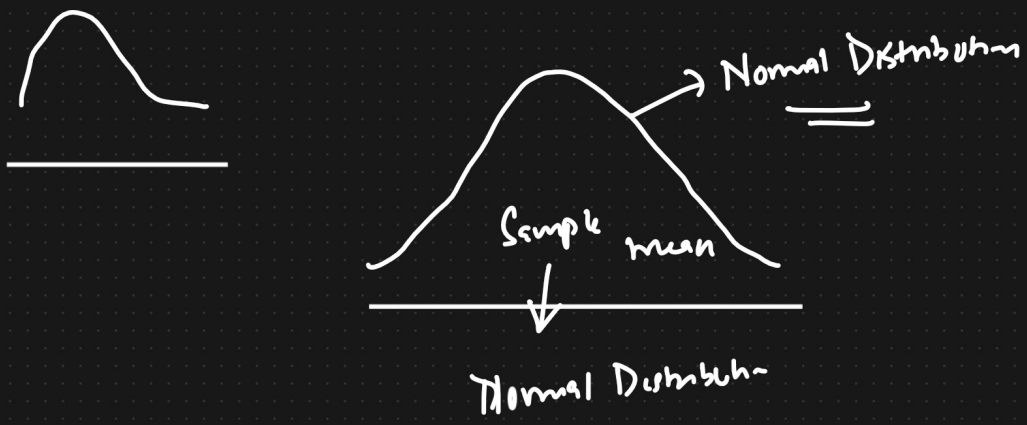
S_1 → $\bar{x}_1$

S_2 → $\bar{x}_2$

} → Sample mean



Now if we take all means and we populate then in form of pdf, then it will get converted into normal distribution



Also read Poisson Distribution

2 Algorithms

① Machine Learning Algorithms ✓
② Deep Learning Algo ✓
③ FLASK & DJANGO ✓
④ Mongodb, Sql ✓
⑤ Blockchain Session ✓