✱ Power law Distribution
or
Pareto Distribution
or
80 - 20 Rule Distribution

organization

80% W — 20% People

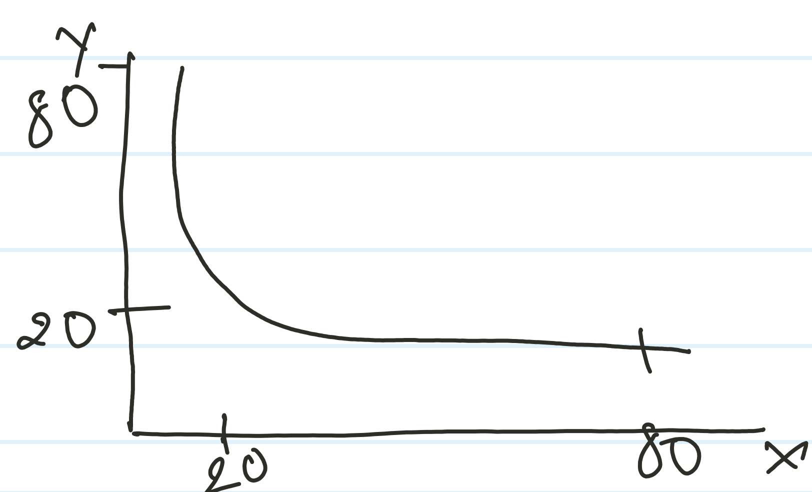
family income

80% — 20%

20% — 80%

Cricket

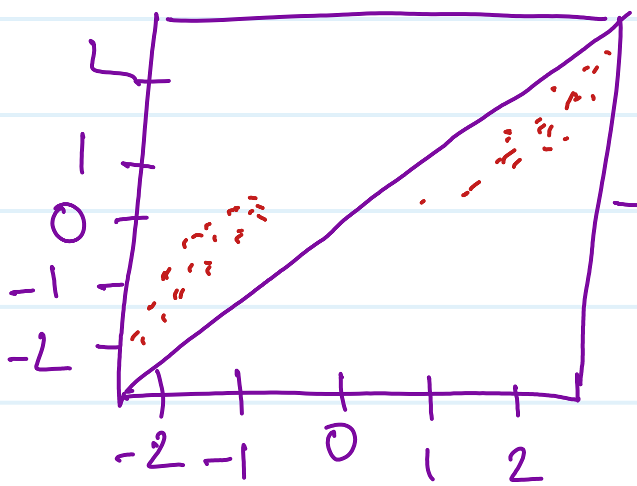
Run	Player
80	20
20	80



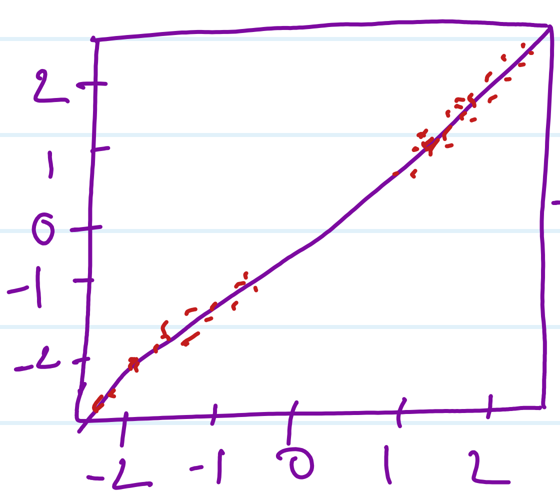
To transform pareto to normal
Dist.

① Box-cox transformation
② lognormal transformation.

☆ Q-Q plot

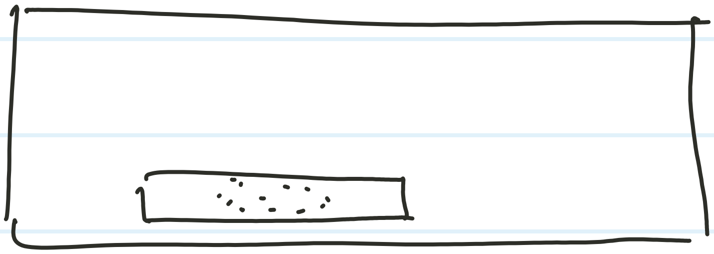


non-Gaussian Dist.



Gaussian Dist.

P- value



Definition - It is the probability for null hypothesis to be true.

Hypothesis testing

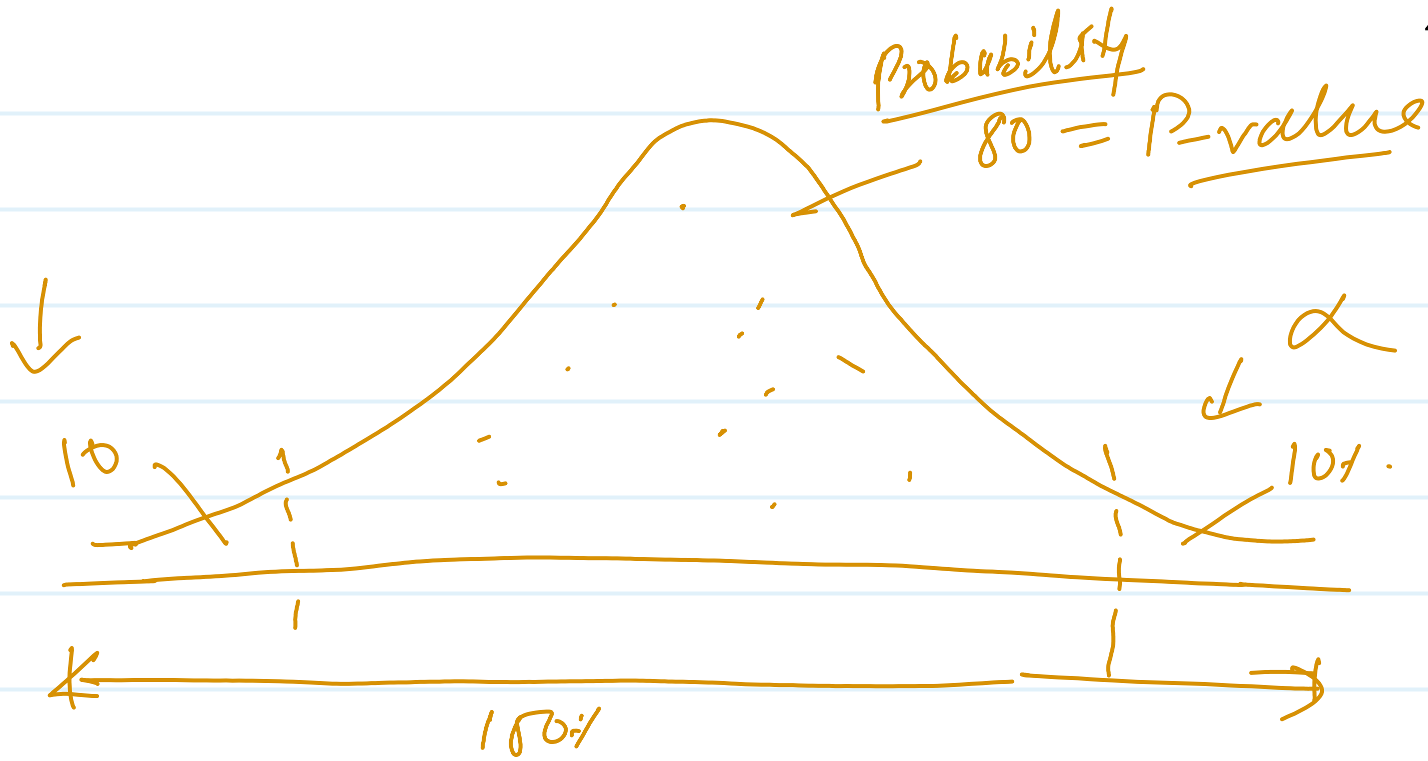
Null hypothesis = H_0

Altenet hypothesis = H_1

Distribution, $\mu = \bar{X} = H_0$

$\mu \neq \bar{X} = H_1$

- $N = 100 \Rightarrow$ 80% relafe
- $n = 10 \Rightarrow$ 60% relef

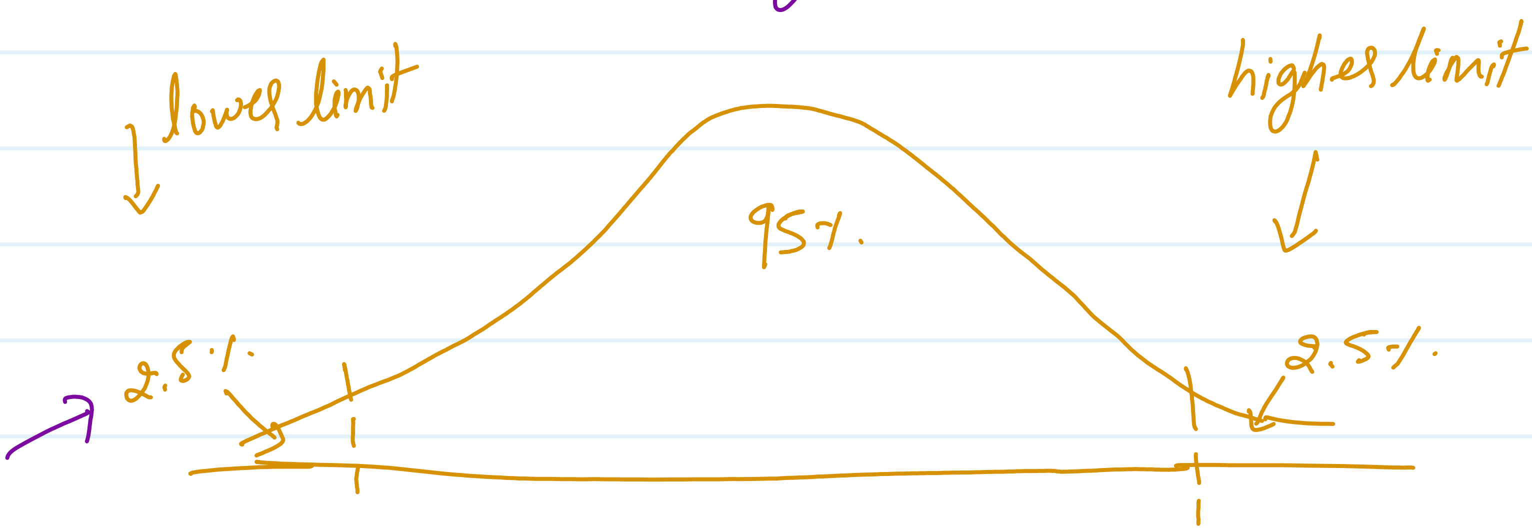


Confidence Intervals About mean

CI

① Point Estimation − $\bar{x}$, μ

Consider for e.g. 95%



Two way to find CI.

① We have given population SD.
② we have not given population SD

E.g.- Suppose SD is given σ
① Avg size of shark in the sea [95%]
we are taking 95% as a CI.
population SD

$$\sigma = 100$$
$$n = 30$$
$$\bar{x} = 500$$

C.I. = point estimate $\pm$ margin of error

$$C.I. = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

α is significant value

$$\alpha = 2.5 + 2.5 = 5\% = 0.05\%$$

$$= 500 \pm Z_{\frac{0.05}{2}} \frac{100}{\sqrt{30}}$$

$$Z_{0.025} = 1 - 0\ 025 = 0.975$$

Z-table $0.975 = 1.96$

Z-table $\boxed{1.9 + 0.06 = 1.96}$

lower limit $= 500 - 1.96 \times \frac{100}{\sqrt{30}}$

$= 464.21$

upper limit $= 500 + 1.96 \times \frac{100}{\sqrt{30}}$

$= 535.83$

Range of the population mean is 464 — 535

$$\overline{X} = 500$$

l 464, 480, 500, 510, 520, 535 u

$\mu = \bar{x}$ — H_0

Note — z-test = when population SD is given / $n \geq 30$

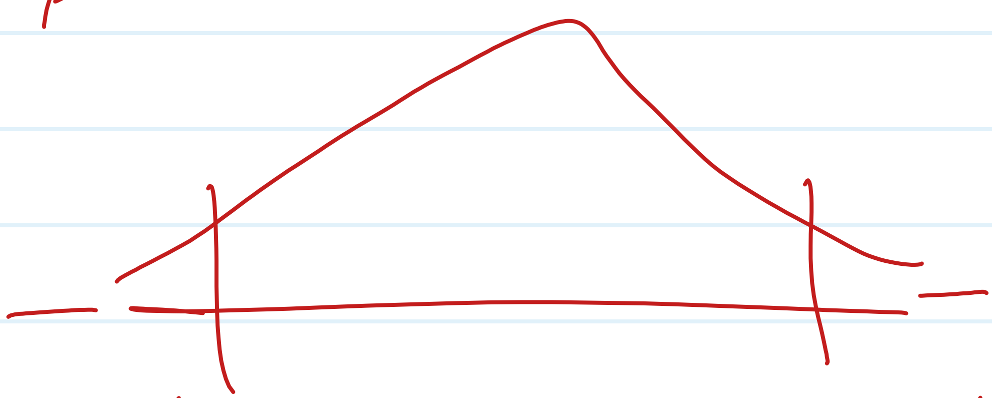
t-test = when population SD isn't given

$n < 30$

~~0.05~~

α = Domain expert

p =



one tail

$\bar{x} > \mu$

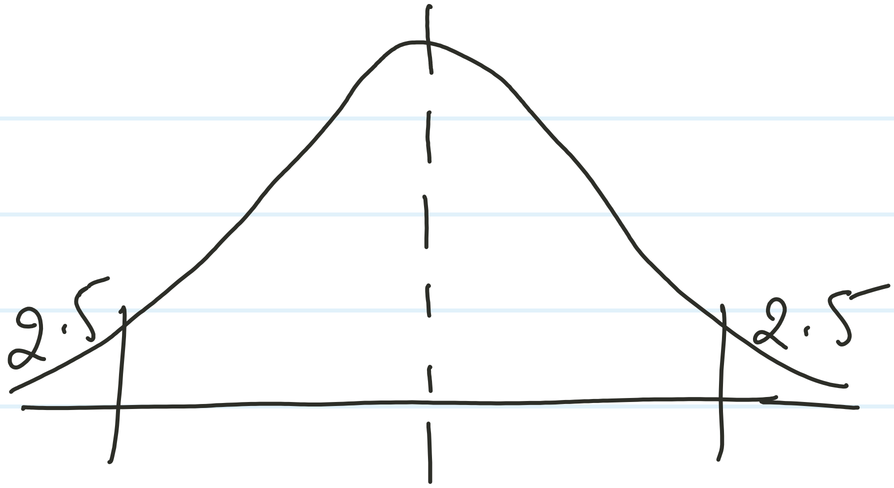
$\bar{x} < \mu$

two tail

$\bar{x} = \mu$

One-Tailed Test	Two-Tailed Test
A test of any statistical hypothesis, where the alternative hypothesis is **one-tailed** either right-tailed or left-tailed.	A test of a statistical hypothesis, where the alternative hypothesis is **two-tailed.**
For one-tailed, we use either **>** or **<** sign for the alternative hypothesis.	For two-tailed, we use **≠** sign for the alternative hypothesis.
When the alternative hypothesis specifies a direction then we use a one-tailed test.	If no direction is given then we will use a two-tailed test.
Critical region lies entirely on either the right side or left side of the sampling distribution.	Critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic.
Here, the Entire **level of significance (α)** i.e. 5% has either in the left tail or right tail.	It splits the **level of significance (α)** into half.
Rejection region is either from the left side or right side of the sampling distribution.	Rejection region is from both sides i.e. left and right of the sampling distribution.
It checks the relation between the variable in a singles direction.	It checks the relation between the variables in any direction.
It is used to check whether the one mean is different from another mean or not.	It is used to check whether the two mean different from one another or not.

E.g



$N = 100$

$\mu = 120$

$\alpha = 5$

$n = 30$

$\bar{x} = 140$

will compare $\mu = \bar{x}$

suppose $C.I. = 95\%$ and $5\% = 0.05/2$

$= 0.025$

Two tail test

$Z\text{-table} = 1 - 0.025 = 0.975$

$= 1.96$

<u>lower</u> $= 140 - 1.96 \times \frac{5}{\sqrt{30}}$

$= 138.21$

$$\text{upper} = 140 + 1.96 \times \frac{5}{\sqrt{30}}$$

$$= 141.78$$

$$138 \;—\; 141$$

$$\downarrow$$

$$140$$

$$\mu = \overline{x}$$

$\Rightarrow$ We reject to fail null hypothesis

one tail

$\alpha = 6$

$CI = 100 - 6 = 94$

$1 - 0.06 = 0.94$

$1 - 0.03 = 0.97 \rightarrow$ Two tail