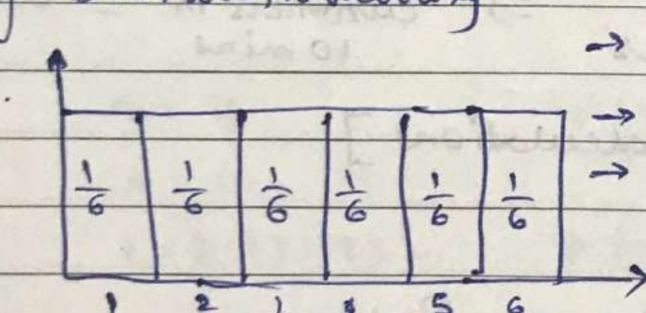
PL 06 - Continuous Probability Distributions

Video 1 : Uniform Distributions : Discrete & Continuous

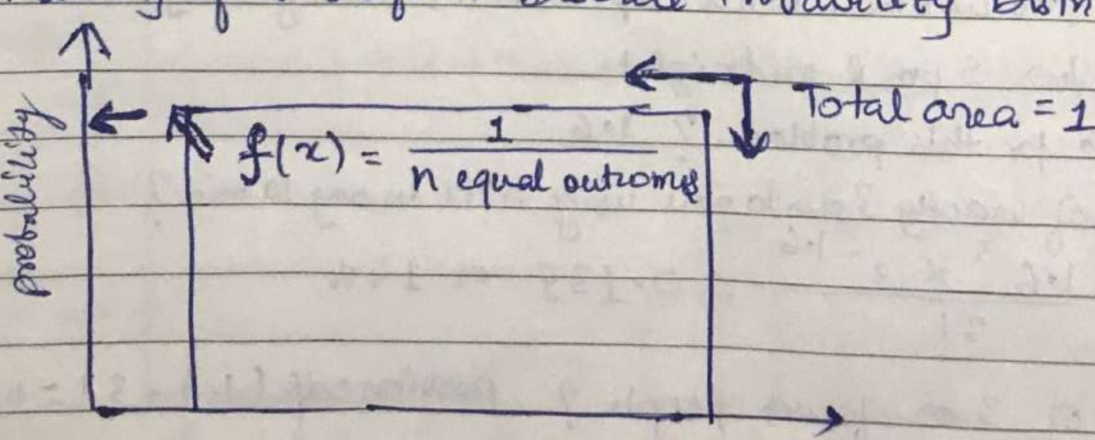
- When dealing with Discrete outcomes, the uniform probability for any specific outcome is defined as $\frac{1}{n}$

- When dealing with continuous outcomes the probability for any specific outcome is undefined.
  $\frac{1}{\infty} \cong 0$: $\lim_{n \to \infty} \frac{1}{n} = 0$
  - Can only find probability over a RANGE of outcomes

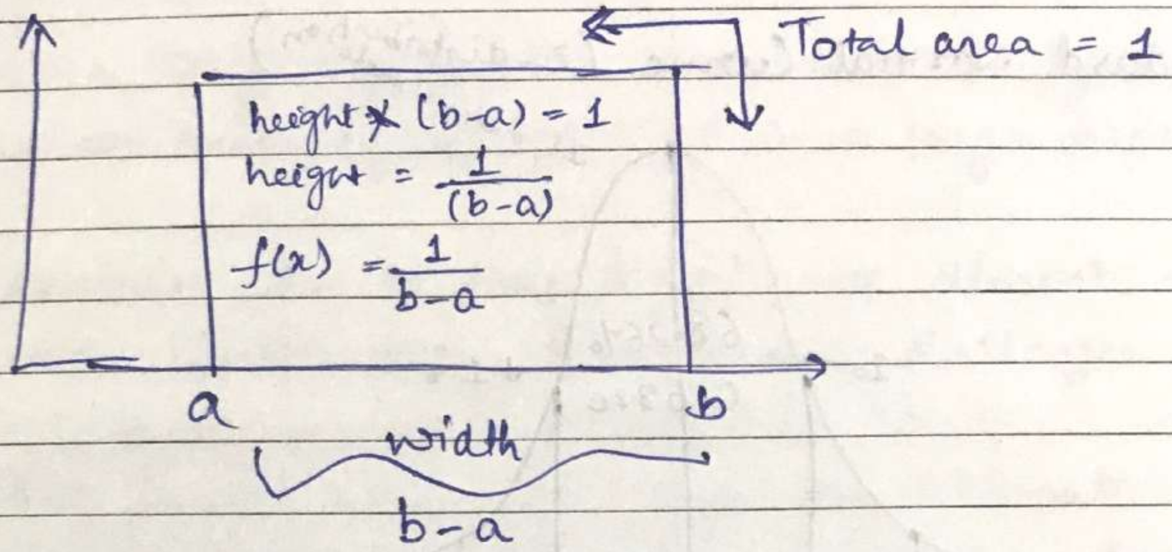
- e.g Die Roll Probability
  → Discrete Outcomes
  → Six outcomes
  → $x = 1, 2, 3, 4, 5, 6$
  → Probability for each outcome is $1/6$



- Anatomy of a Uniform Discrete Probability Distribution



- Anatomy of a Uniform Continuous Probability Distribution.



Expected Value (x) $= \dfrac{a+b}{2}$

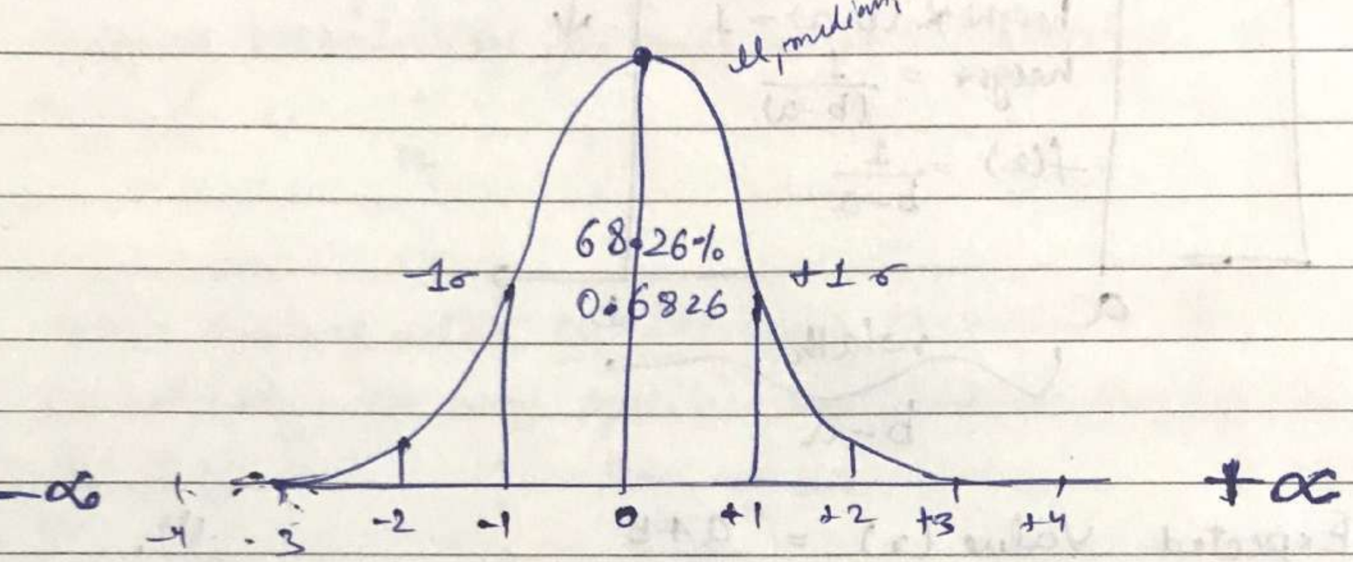
Variance (x) $= \sigma^2 = \dfrac{(b-a)^2}{12}$

Dr-natty.

Standard Deviation (x) $= \sigma = \dfrac{(b-a)}{\sqrt{12}}$

Video 2 : The Probability Distribution : Normal Distribution

- Standard Normal Curve (Z-distribution)



$$P(-1 \leq Z \leq 1) = 0.6826$$
$$P(-2 \leq Z \leq 2) = 0.9544$$
$$P(-3 \leq Z \leq 3) = 0.9974$$

→ This is a specific type of distribution called Z-distribution. We standardize our distribution by making $\mu = 0$ & $\sigma = 1$

→ Area under the curve = 1

→ Cumulative probability : $-\infty \leq Z \leq 0 = 0.5$

→ NORM.DIST in excel

→ σ defines the shape of distribution
→ μ, median, defines the slides distribution side-to-side.

Video 3 : The Probability Distribution : To Z or to T ?

- When doing quantitative research or analysis, we are most often interested in a large population.

- However, due to time & cost, we almost always use Sample data to represent the Larger population

- But Sample Data is always an estimate or approximation of the larger population from which it is selected.

- As Sample Size n becomes small ↓, there is less chance ↓ of representing entire population.

- Additionally, we do not know anything about our population - i.e its mean, variance, s.d.

- Larger Sample, more likely to capture Natural Variation.
- Small sample, increases the chance that either we miss variation or over-emphasize it.

- (+) There is a point when increasing sample size offers no more statistical benefits.

- ## t - Distribution(s)

- When sample size is $n \leq 30$ and/or we do not know the variance/standard deviation of the population, we use t-distribution instead of z-distribution.

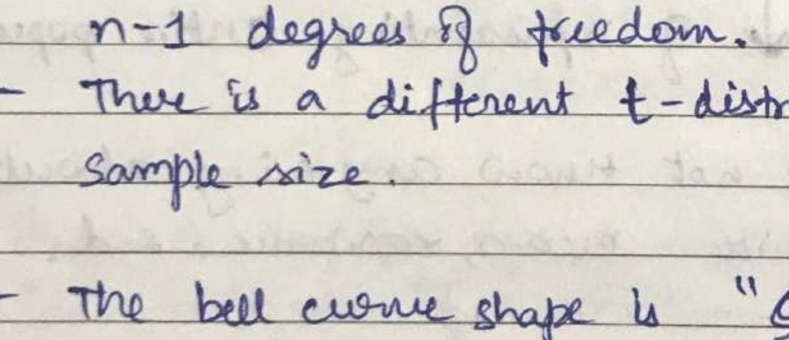
-- The t-distribution allows us to use small samples $n \leq 30$

-- Margin of error is little bit wider.

-- It takes sample size into account using $n-1$ degrees of freedom.
-- There is a different t-distribution of any given sample size.

-- The bell curve shape is "SQUISHED" in the middle and "fatter" on the ends (tails)
Squisher & fatter, the smaller the ~~smaller~~ sample size.

-- as $n > 30$ and definitely $n \geq 100$, the t-distribution is same as z-distribution.

--


-- The probability of having a value farther from the true mean is greater when the sample size is small ; fatter tails.

or

-- Greater uncertainty on tails due to small sample sizes

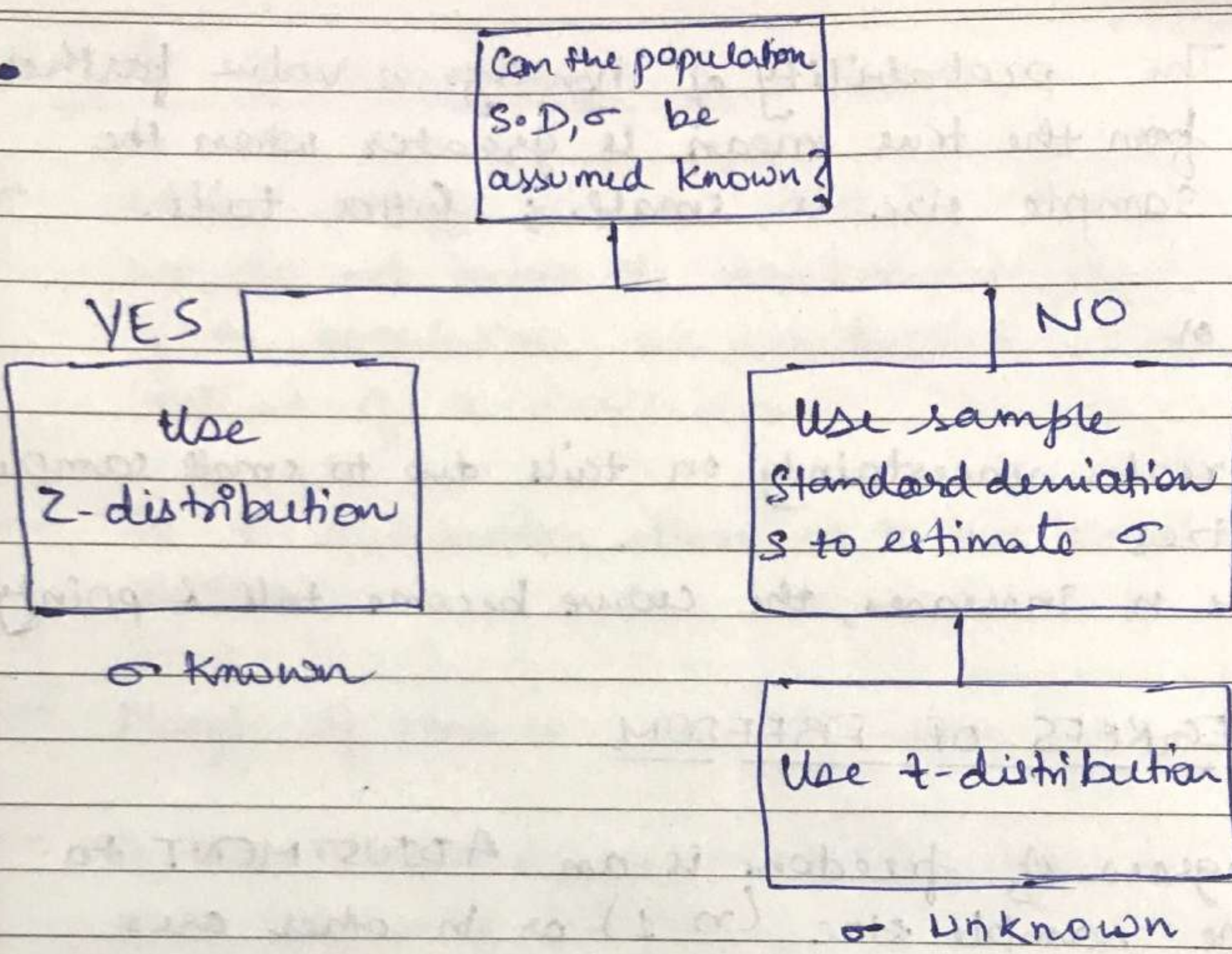
As n increases, the curve become tall & pointy.

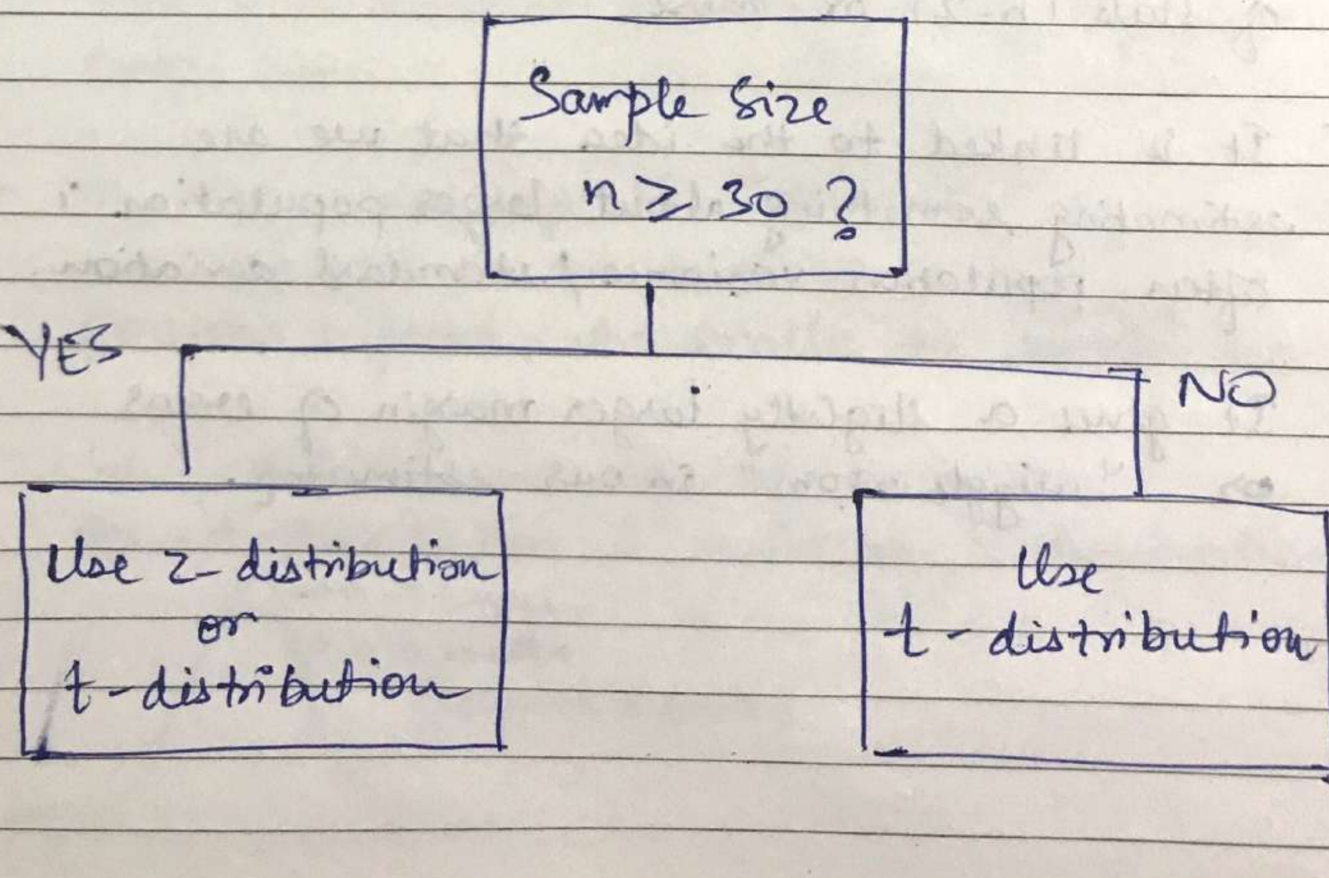
- DEGREES OF FREEDOM

-- Degrees of freedom is an ADJUSTMENT to the sample size (n-1) or in other ares of stats (n-2) or more.

-- It is linked to the idea that we are estimating something about larger population ; often population variance / standard deviation.

-- It gives a slightly larger margin of error or "wiggle room" in our estimates.

- 
Can the population S.D, σ be assumed Known?
YES
NO
Use z-distribution
σ known
Use sample standard deviation s to estimate σ
Use t-distribution
σ unknown

- 
Sample size n ≥ 30 ?
YES
NO
Use z-distribution or t-distribution
Use t-distribution

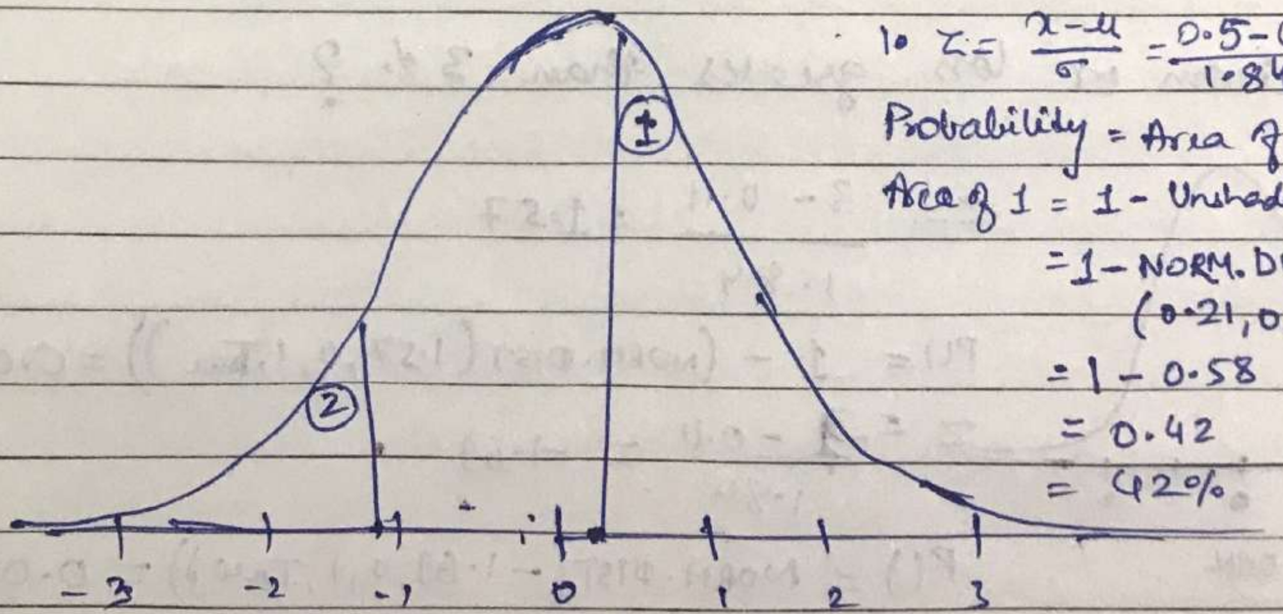
Video 4: Probability Distributions: NORMAL DISTRIBUTION

- The Standard Normal Curve - Z distribution
  $\mu = 0$
  $\sigma = 1$

- NORM.DIST (x , mean , standard dev, cum)
  - x ↓ upper bound of the cum distribution
  - mean ↓ = 0 b'coz, we are using standard normal curve
  - standard dev ↓ = 1
  - cum ↓ True

- e.g. Apple Daily Returns
  - $\mu = 0.11\%$
  - $\sigma = 1.84\%$

1. What is the Probability, for any day, of a return greather than 0.5
2. " " " " " " ...., of a loss greater than 2% ?
3. " " " " " " " ", of a return b/w 0% & 1% ?
4. " " " " " " " " ", of a return OR loss greater than 3%.



1. $Z = \frac{x-\mu}{\sigma} = \frac{0.5-0.11}{1.84} = .21$

Probability = Area of 1

Area of 1 = 1 - Unshaded area

= 1 - NORM.DIST (0.21, 0, 1, True)

= 1 - 0.58

= 0.42

= 42%

2. loss greater than 2%?

$$z = \frac{x-\mu}{\sigma} = \frac{-2-0.11}{1.84} = -1.15$$

Probability = NORM.DIST(-1.15, 0, 1, True) = .125

⇒ Probability = 12.5%

3. Return between 0% and 1%?

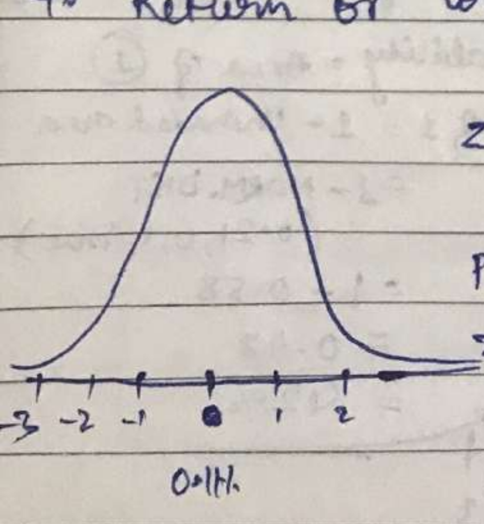
→

→ For 1%, $z = \frac{1-0.11}{1.84} = 0.48$ | P(1%) = 0.69

→ For 0%, $z = \frac{0-0.11}{1.84} = -0.06$ | P(0%) = 0.48

→ Req Probability = 0.69 − 0.48 = 0.21

4. Return or loss greater than 3%?



$$z = \frac{3-0.11}{1.84} = 1.57$$

P() = 1 − (NORM.DIST(1.57, 0, 1, True)) = 0.058

$$z = \frac{-3-0.11}{1.84} = -1.69$$

P() = NORM.DIST(−1.69, 0, 1, True)) = 0.046

Total Probability = .104 = 10.4%

- Apple has a slightly higher mean daily return.
- Apple also has more variation, a wider distribution with "fatter" tails.
- Ge offers slightly lower mean daily return,
- With much less variation.

Video 5 : Probability Distribution: Is my Data NORMAL?

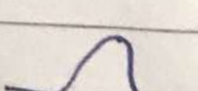
(i) Look at your Data Graphically first., before starting any analysis.

(ii) Know data, look for patterns, initial relationships. Our data may have EXCESS SKEW (lopsided), kurtosis (very fat tails), bi-modal (two humps) or a distribution other than normal distribution.

(iii) These 5 tools for pre-liminary analysis to know whether the data fits for normal distribution.

- Histograms
- Stem & Leaf Plots
- Box plots (Box & whisker Plots)
- P-P Plots
- Q-Q Plots

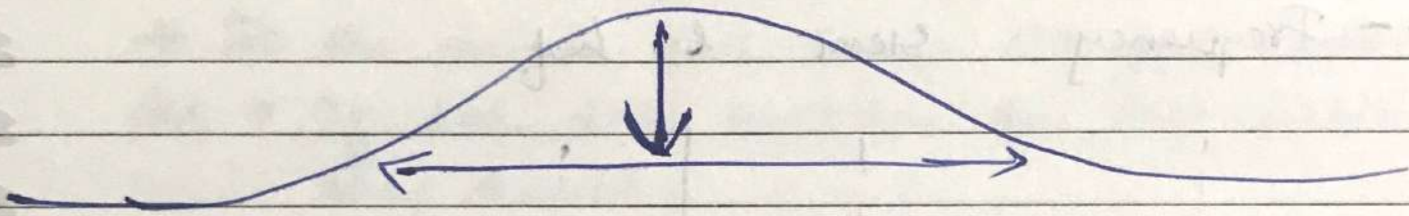
(iv) the main reason to check for normality, because many statistical techniques assume that data fits a normal distribution.

- How can we tell if our data fits this shape? 
- Does our data have "goodness of fit" relative to the normal distribution?

- EXCESS KURTOSIS

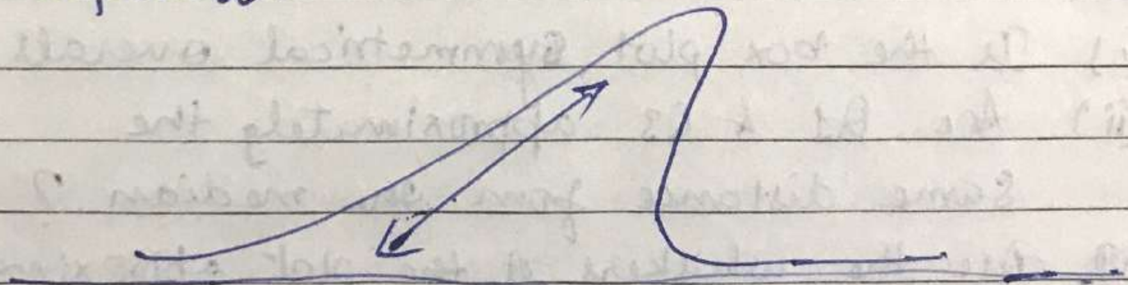
-- More probability than excpected in the tails of the distribution due to extreme values away from mean.

-- Provability (values) are pushed away from the mean toward the tails.



- EXCESS SKEWNESS

-- More probability than expected is on one side of the distribution versus the other; lopsided.



- Oftentimes data fits another type of distribution

Lognormal Exponential

Uniform Weibull

- Histograms

-- can be misleading. The look of a histogram is largely dependent on the "bin" size.

- Stem & Leaf Plots

-- Sideways histogram

-- Frequency   Stem & Leaf

- Box Plot

Look for:

(i) Is the box plot symmetrical overall?

(ii) Are Q1 & Q3 approximately the same distance from the median?

(iii) Are the whiskers of the plot approximately the same length?

- P-P Plots

-- In a P-P Plot we compare the cumulative probability of our empirical data with an ideal 'test' distribution. Let's say, the normal distribution.

-- Questions to ask.

→ Do the points fall in a straight line?
Ans → If our data matches the test distribution they should.

- Q-Q Plot

-- In a Q-Q Plot, we compare the Quintiles of our empirical data with the ideal.

- Important characteristics depicted in the video. Refer PL06 - video 5 for it.