VII Histogram and Skewness :-

Remember to learn : 5 no. summary, Box Plot

6.17 Histogram :-

Histogram are very useful in descriptive statistics to summarize the data and take out information regarding distribution of data.

Bins in Histogram Bin refers to no. of intervals or buckets for a particular no. of datasets

ex:-

for dataset Ages = {10, 12, 14, 18, 24, 26, 30, 33, 36, 37, 40, 41, 42, 50, 51}

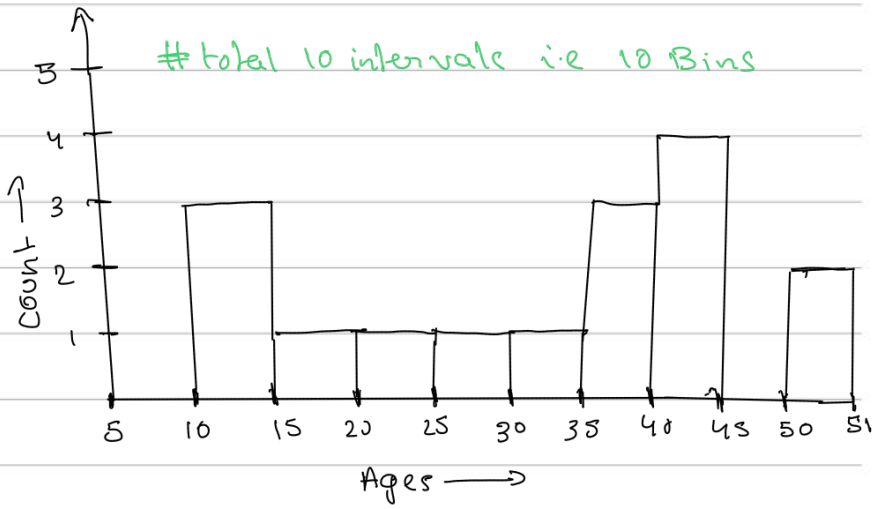
i.e.,

Calculating Ages difference for bin size = 10

$$\text{min data range} = \frac{\text{max} - \text{min}}{\text{Bin Size}} = \frac{51 - 0}{10} = 5.1 \approx 5$$

So,

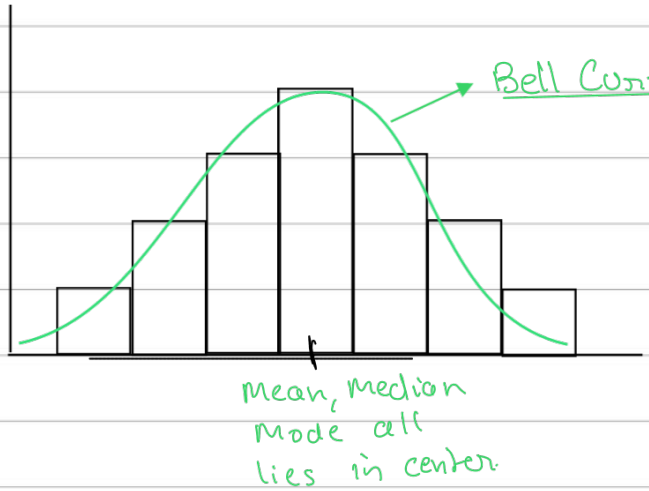
Histogram :-



Now, smoothening the above Histogram we get,

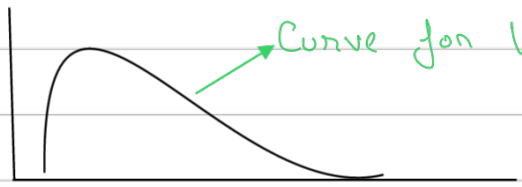
Probability Distribution Function (PDF):- This basically talks about how my data is getting distributed. PDF is achieved by smoothening the Histogram by a concept called as Kernal Density Estimator (KDE)

6.1.2Y Bell Curve



Bell Curve is a type of Probability distribution function for which all central tendencies (like mean, median & mode lies at the center of the distribution and such type of distribution is cfa "Normal / Gaussian Distribution"

6.1.3Y Log Normal Distribution Curve

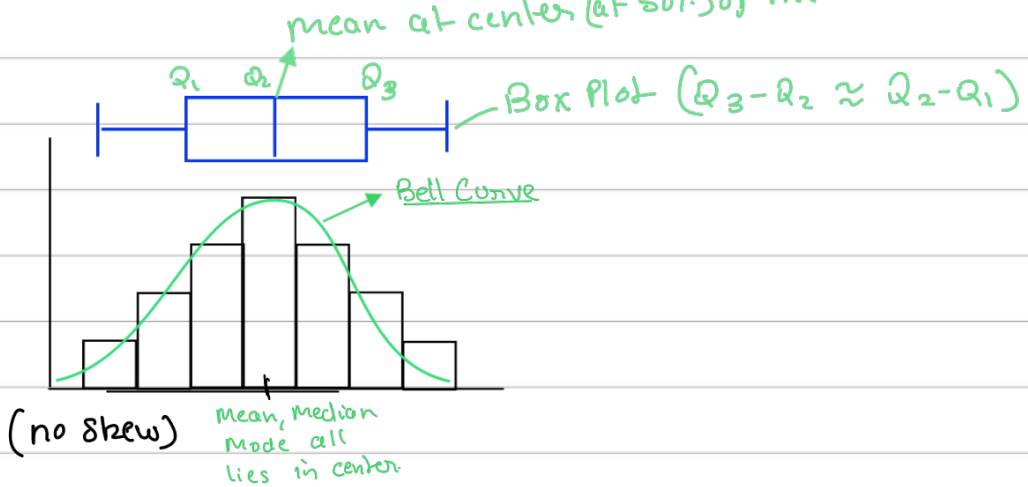


6.2Y Skewness :-

6.2.1Y No Skewness :-

Normal Distribution also known as Gaussian Distribution is a Bell shaped symmetrical distribution in which the mean, median and mode all are perfectly at the center (i.e all are

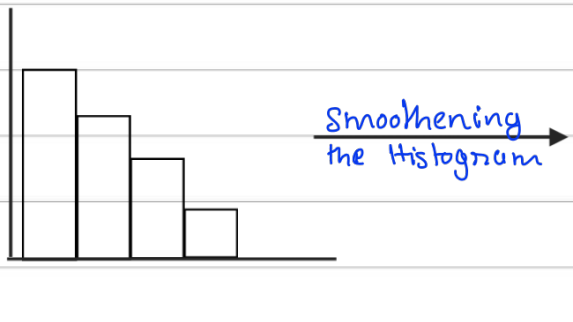
equal). For such kind of distribution skewness is zero

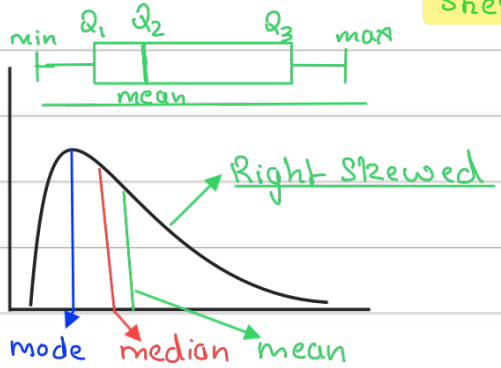


6.2.2) Right Skewed Distribution :- Log Normal Distribution

also c/a Positive Skewed

(Q3-Q2 > Q2-Q1) for Right Skew





∴, For Right Skew distribution
mean > median > mode

Q) Why mean > median > mode for Right Skew Distribution?

Sol^n ->

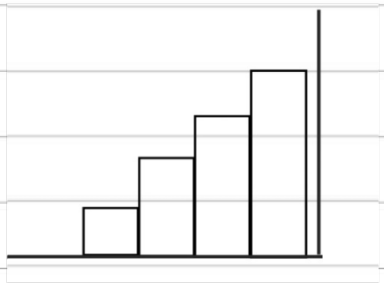
Mean: as for Right Skewed type of distribution the maximum no. of distribution or data is on the right so mean is shifted most to the right

Median :- As we know in case of an overlay median shift is alway less than the mean ∴, Median is shift is less that mean

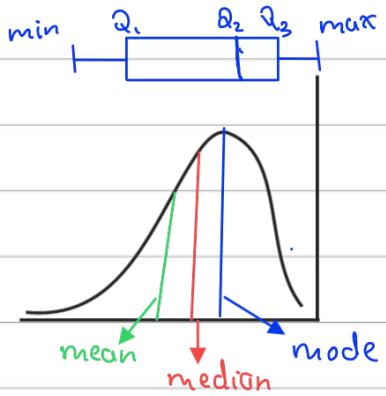
Mode :- As mode is always at the highest freq ∴, it will found at the maxima.

6.2.3.4 Left Skewed Distribution :-

$Q_2 - Q_1 > Q_3 - Q_2$ for Left Skew



Smothening →



∴,