

# Histogram And Skewness

A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable and is used to visualize the shape, central tendency, and variability of a dataset.

Ages = { 11, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50 }  $\Rightarrow$  Histogram

0-50

① No. of bins = 10

$$\frac{50}{10} = 5 \rightarrow \text{bin size}$$

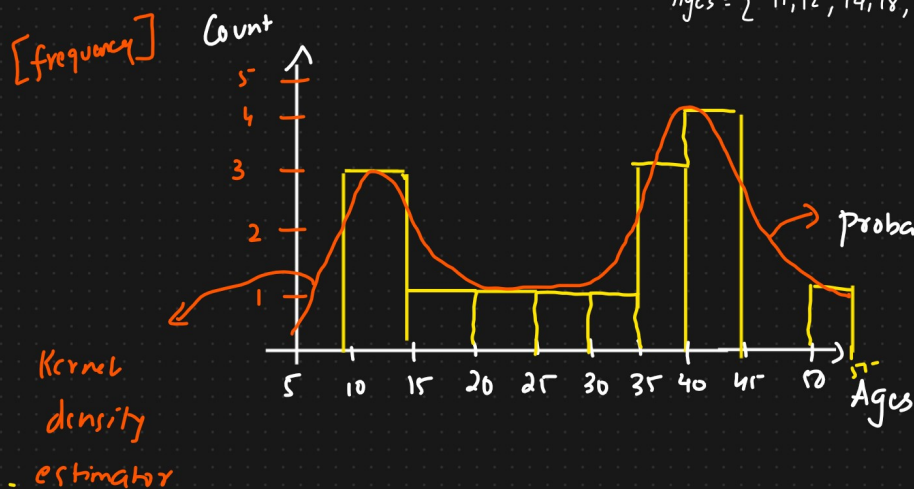
Bins  $\rightarrow$  [ 0-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50 ]

Ages = { 11, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50 }

$\geq 25$   $< 30$

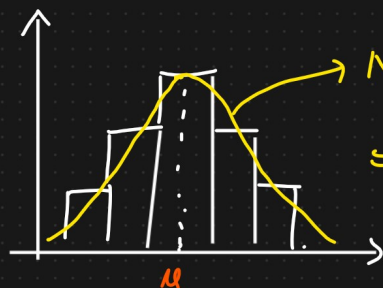
Distribution of Numerical DATA

probability distribution of a continuous variable



Ages = { - - - - - }

## Skewness



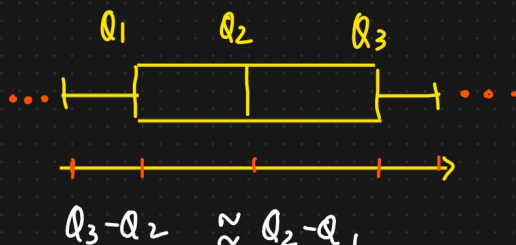
$\Rightarrow$  Symmetrical Distribution  $\Rightarrow$  No Skewness



mean, median, mode

The mean, median and mode are all perfectly at the center.

## Box plot

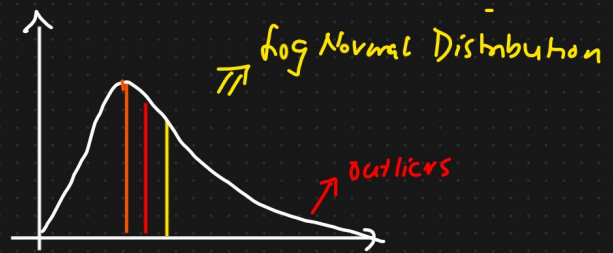


$$\boxed{\text{Mean} = \text{Median} = \text{Mode}}$$

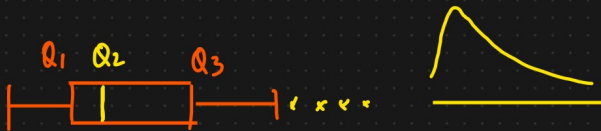
## ② Right Skewed



$\Rightarrow$  Positive Skewed  $\Rightarrow$



Box plot



$$Q_3 - Q_2 > Q_2 - Q_1$$

$$\boxed{\text{mean} > \text{median} > \text{mode}}$$

## ③ Left Skewed Distribution



$\Rightarrow$  Negative Skewed

Box plot



$$Q_2 - Q_1 > Q_3 - Q_2$$

Relationship

$$\text{mean} \leq \text{median} \leq \text{mode}$$