

Statistics

① Descriptive

① Aggregate Measure

min(), max(), sum()

count(), average()

② Central Tendency Measure

Mean, Median, Mode

↓
No outliers

↓
Outliers

↓
Categorical
repetitive

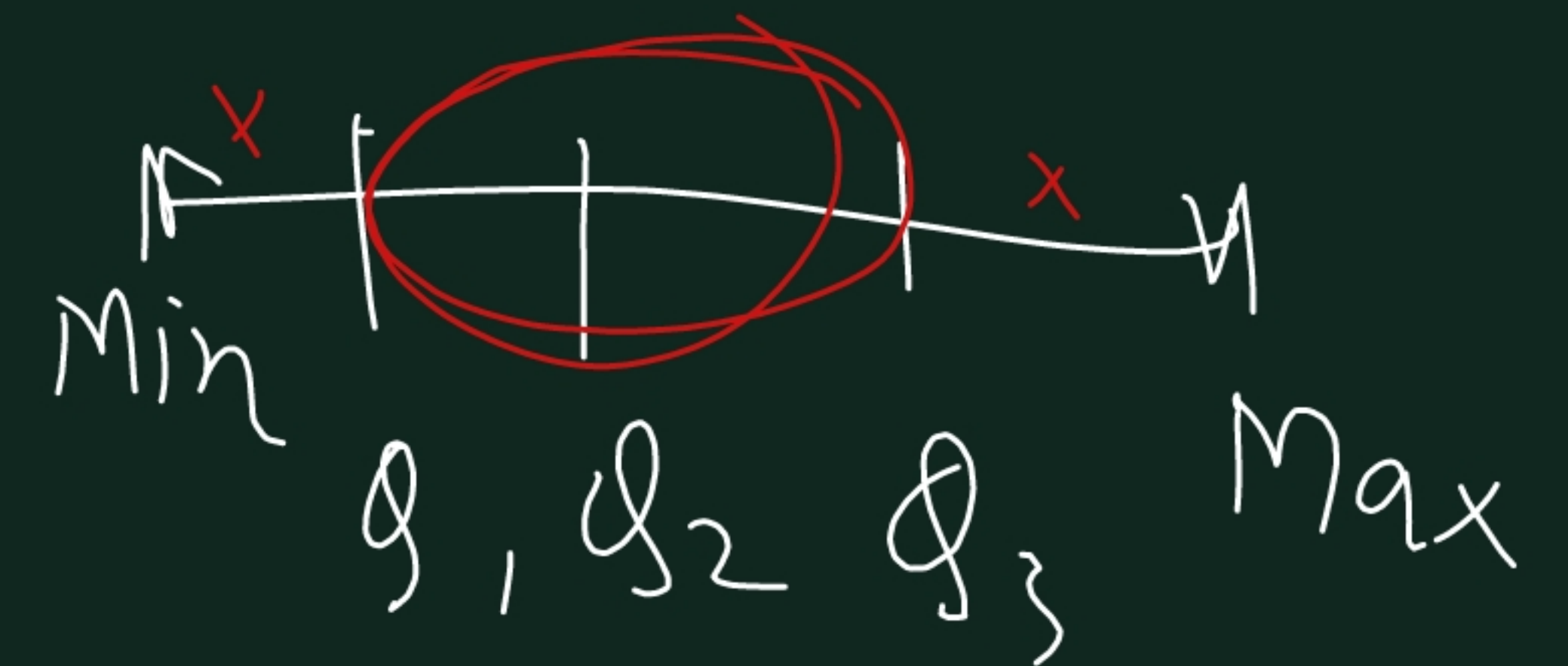
② Spreadness Measure

(i) Range = Max - Min



When min & max value same, fail

(ii) Interquartile Range = $Q_3 - Q_1$



⊕ Outlier Detection Using IQR :

Range & IQR fails, becoz they do not consider central tendency.

$$L.B = Q_1 - 1.5 * IQR$$

$$U.B = Q_3 + 1.5 * IQR$$

⇒ High range/IQR value → Higher spreadness

When Q_3, Q_1 is same, IQR fails.

✓ (iii) Variance = distance of each datapoint from mean

$$var = \sum_{i=0}^n \frac{(d_1 - \mu)^2 + (d_2 - \mu)^2 + \dots}{N}$$

$$var = (value)^2$$

to convert this into similar units like data

$$(iv) STD = \sqrt{var} \\ = \underline{value}$$

$$Var = (32)^2 \text{ m}^2$$

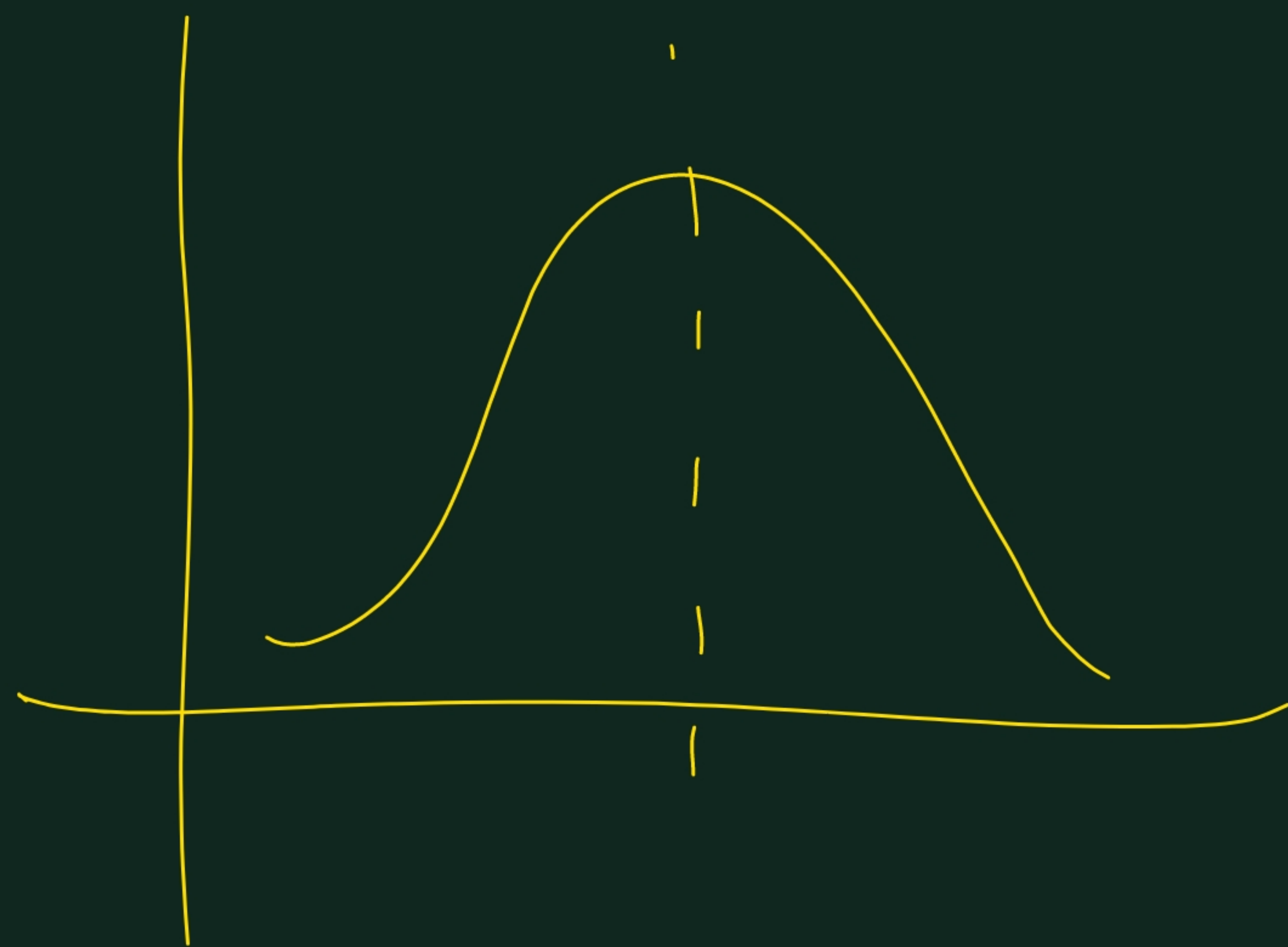
$$Std = 32 \text{ m}$$

Shape Measure

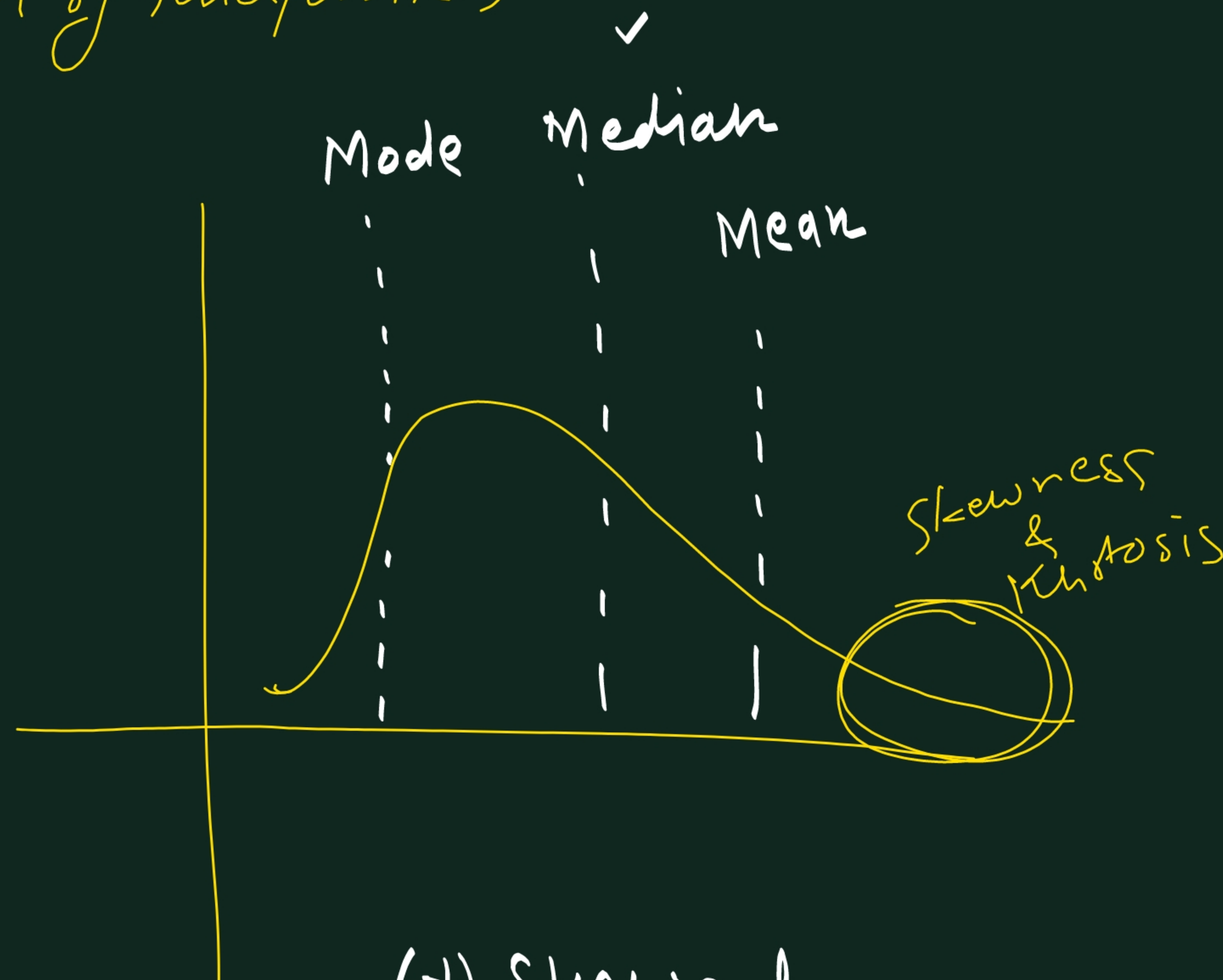
(i) Skewness \div Direction of outliers

(ii) Kurtosis \div Strength of tail/outliers

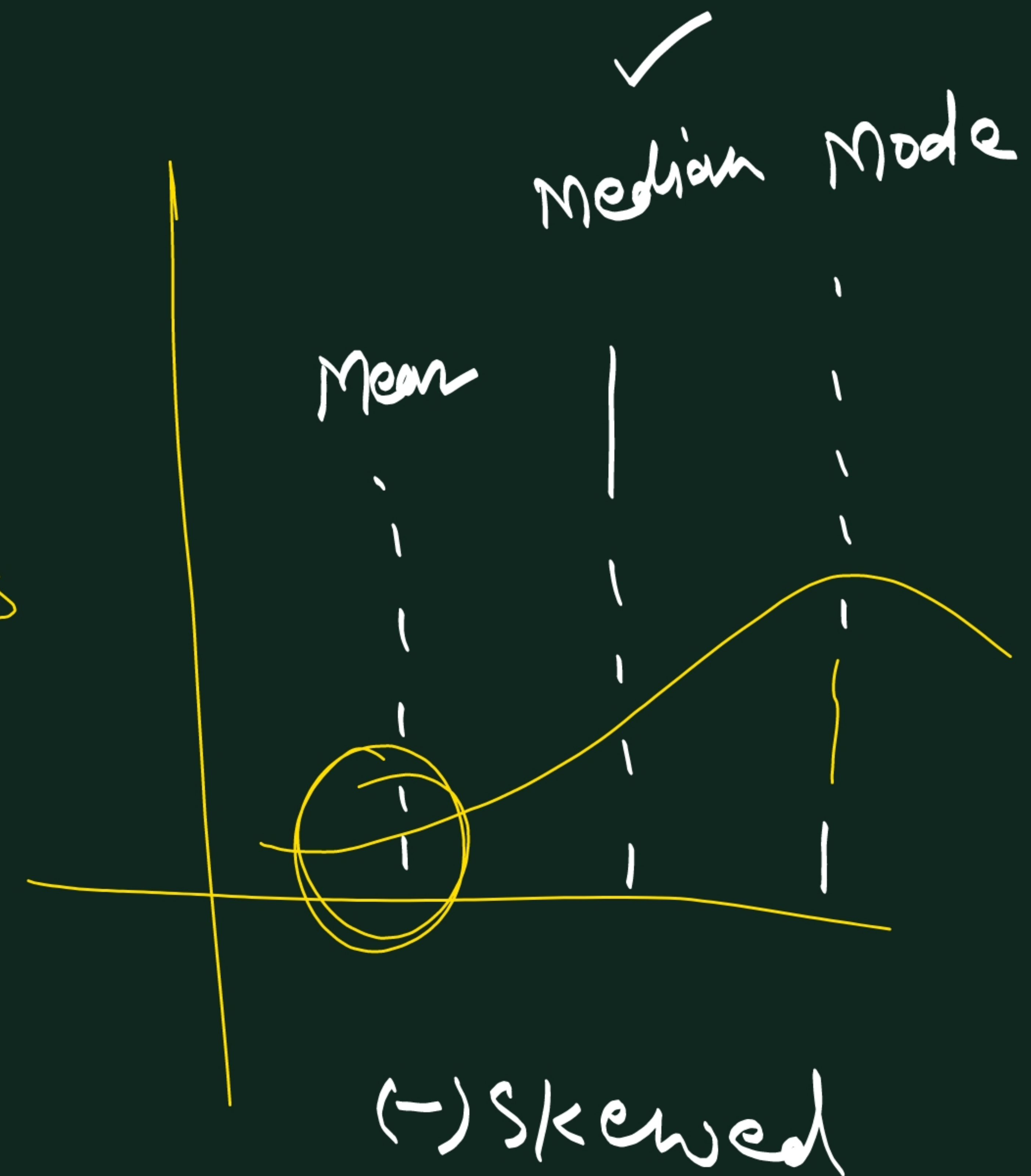
Mean = Median = Mode



Normal



(+) Skewed



(-) Skewed

