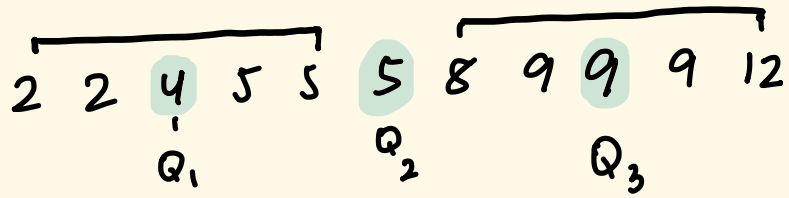


Always arrange the numbers in ascending order first

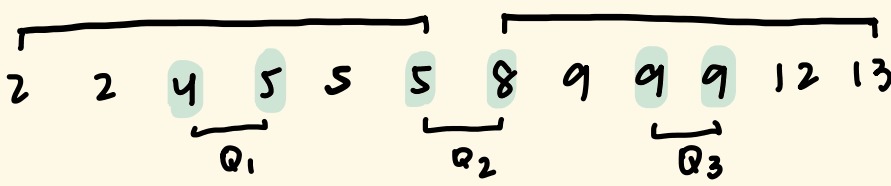
Quartiles

Range	Largest # - smallest # - measures the spread of data within the entire distribution
Interquartile range (IQR)	- $Q_3 - Q_1$ - The spread of data around the median of the distribution (consistency)

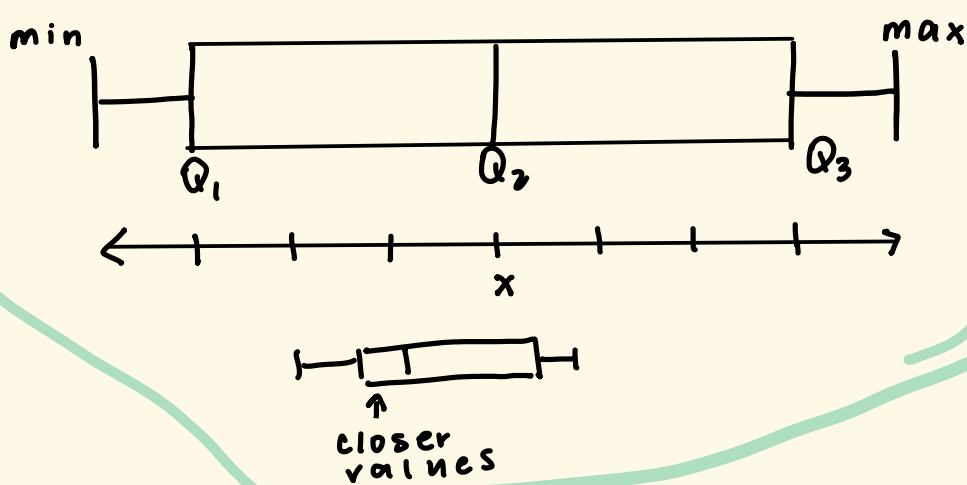
Odd # of values



Even # of values



Box and Whisker Plot



Statistics

Mode	Most repeated number
Median	The middle number
Mean	The average

Discrete data	- All whole numbers (1, 5)
Continuous data	- Have decimals (42, 64.5)

Mean if all values are multiplied by constant $k = (a)(\text{mean})$
Mean if all values are increase by a constant $a = a + \text{mean}$

Standard Deviation

- Measures the spread of data around the mean
- Larger means wider spread

for grouped data:

$$\sigma = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}} = \sqrt{\frac{\sum fx^2}{\sum f} - (\bar{x})^2}$$

use 5sf to be accurate

Ungrouped data

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \text{ or } \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

Ex: 1, 2, 3, 4, 5

$$\text{mean} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3$$

$$\sigma = \sqrt{\frac{1^2+2^2+3^2+4^2+5^2}{5} - (3)^2}$$

$$= \sqrt{11 - 9}$$

$$= \sqrt{2} = 1.41 \text{ (3sf)}$$

SD if all values are multiplied by constant $k = |k|(\text{SD})$
SD if all values are increase by a constant = no change

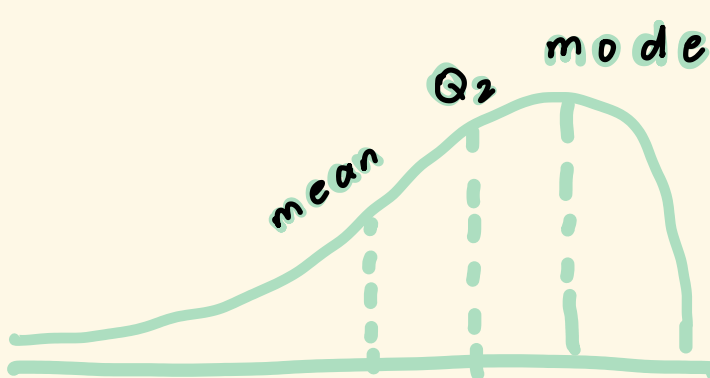
- Standard deviation remains the same if:
- 1) All values of a data set are shifted laterally
 - 2) The graph is laterally inverted
- Symmetrical = smaller SD

① Symmetrical



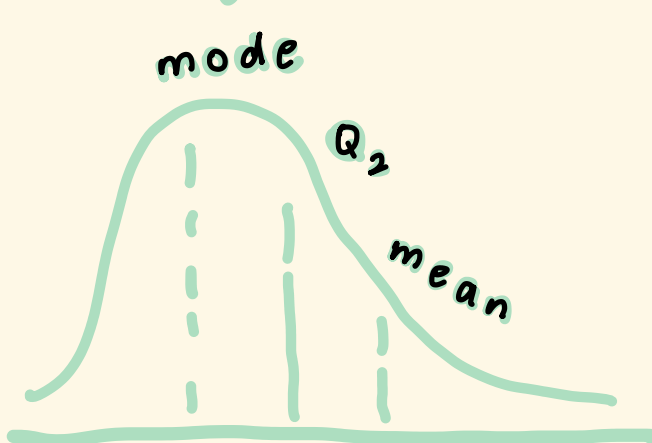
- The mean, median, and mode would be around the same

② Left skewed



- Long left tail
- Mean < median < mode

③ Right skewed



- Long right tail
- Mode < median < mean

Percentiles

Split data

$$P_{25} = Q_1 = T(\frac{1}{4}n)$$

$$P_{50} = Q_2 = \text{median}$$

$$P_{75} = Q_3 = T(\frac{3}{4}n)$$

or ex:

$P_{50} = 50\% = 50\%$ of the distribution is less than n

Culminative Frequencies

Number of hours	Frequency
$0 \leq t \leq 2$	3
$2 \leq t \leq 4$	5
$4 \leq t \leq 6$	16
$6 \leq t \leq 8$	12
$8 \leq t \leq 10$	4

Number of hours	Frequency
$t \leq 2$	$0+3=3$
$t \leq 4$	$3+5=8$
$t \leq 6$	$8+16=24$
$t \leq 8$	$24+12=36$
$t \leq 10$	$36+4=40$

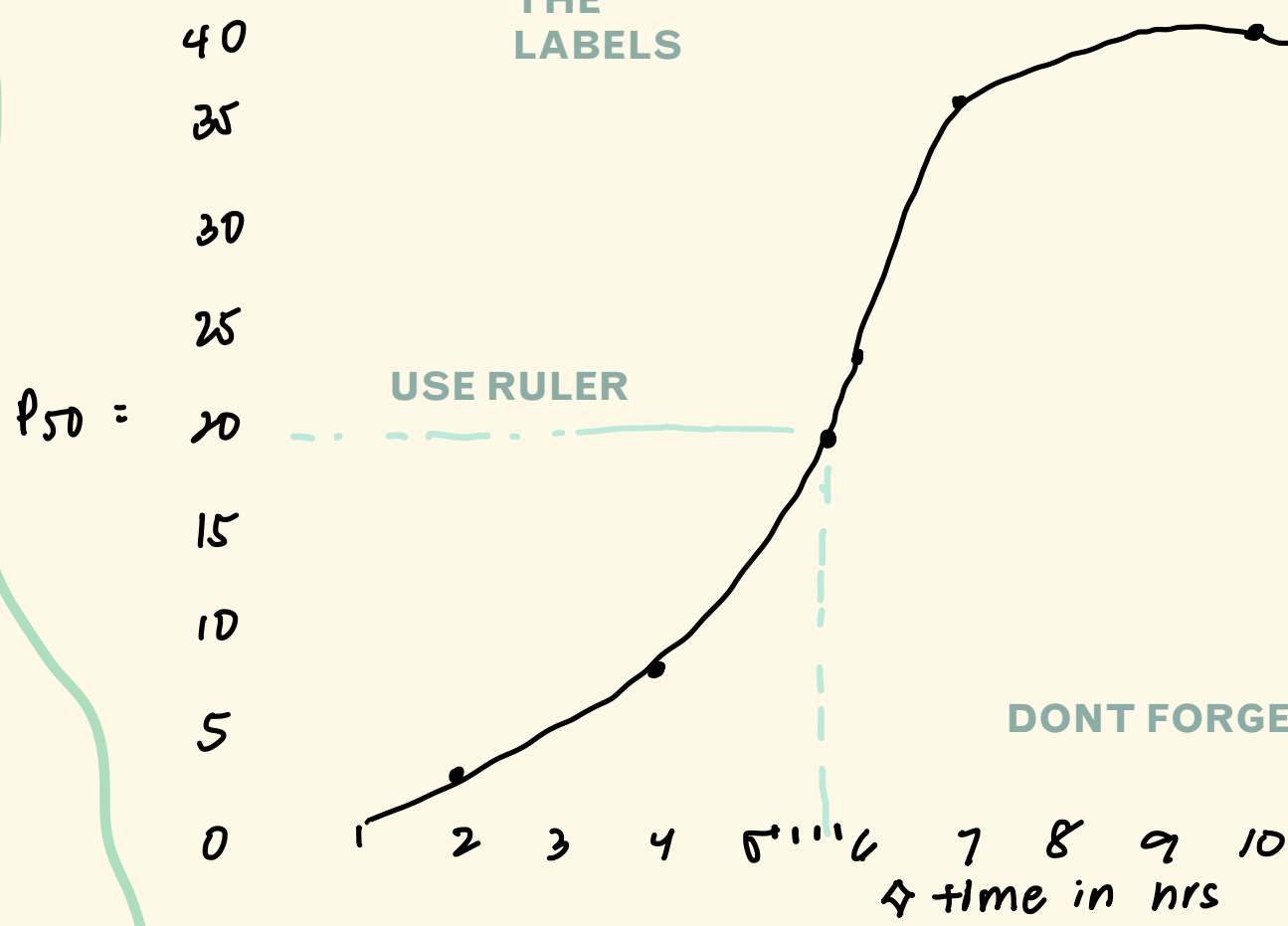
Total Frequency = the size of the data base

Total # of values

plotting:

- (2, 3)
- (4, 8)
- (6, 24)
- (8, 36)
- (10, 40)

culminative frequency



DON'T FORGET TO WRITE THE LABELS

Potential issues in Statistical Analysis (self reading)

1. Sample size too small \neq generalize
2. 2 different groups with large difference in size (not accurate)
3. Graphs have to be of the same scale