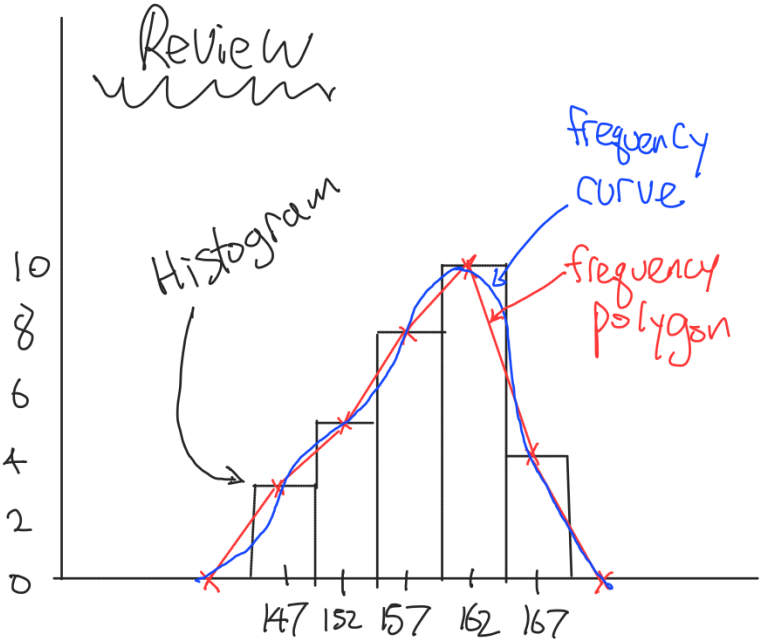
# Maths 22 Measure of Dispersion

Review



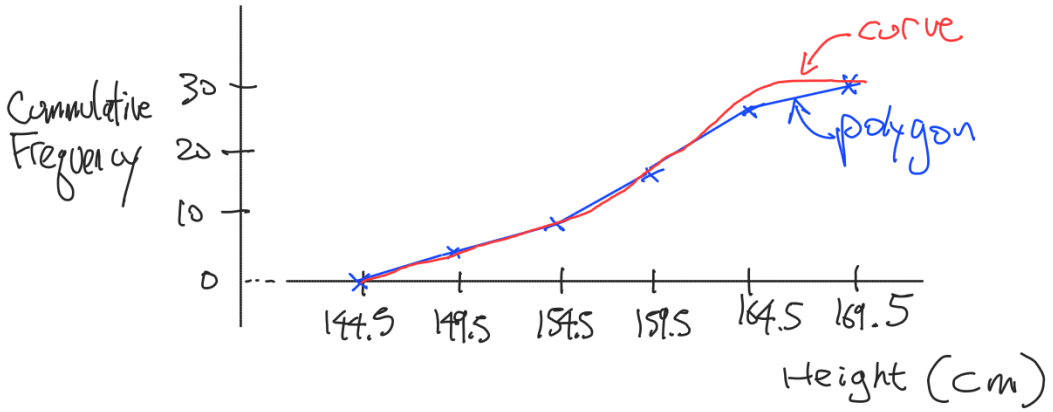

**Histogram**

**Frequency Curve/Polygon**

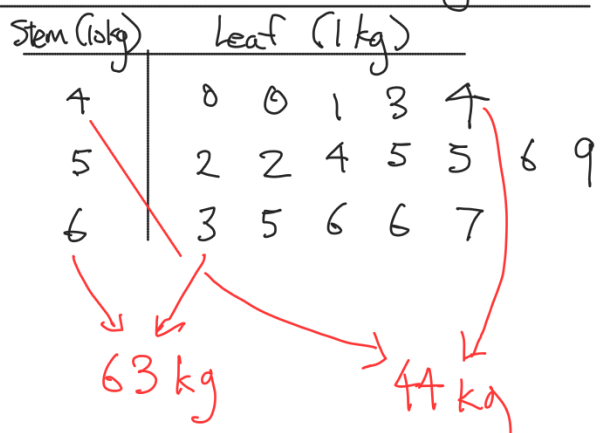
represents the distribution of a set of data

## Cummulative Frequency Polygon Curve

| Height less than | Cumulative Frequency (curve/polygon) |
|---|---|
| 144.5 cm | 0 |
| 149.5 cm | 3 |
| 154.5 cm | 8 (=3+5) |
| 159.5 cm | 16 (=3+5+8) |
| 164.5 cm | 26 (=3+5+8+10) |
| 169.5 cm | 30 (=3+5+8+10+4) |



# Stem and Leaf Diagram




# Mean, Median and Mode

平均值 中間值 多數值

**mean:** $\bar{x} = \dfrac{x_1 + x_2 + \cdots x_n}{n}$ or $\bar{x} = \dfrac{f_1 x_1 + f_2 x_2 + \cdots f_n x_n}{f_1 + f_2 + \cdots f_n}$

**median:** median = { middle datum (for odd no. of data); average of 2 middle data (for even no. of data) }

median = $\frac{n}{2}$th data read from cumulative freq data

**mode:** mode = most frequent data value (ungrouped)

mode = frequency (f) is highest (grouped)

# Range and Inter-quartile Range

Consider 2 boxes of oranges

| Box P | Box Q |
|---|---|
| 155, 155, 155 | 140, 148, 152 |
| 161, 164, 182 | 164, 178, 190 |

Dispersion: Box Q more dispersed: Box P ⇒ 155 → 182; Box Q ⇒ 140 → 190

# Range

Range = Largest value − Smallest value

Box P range = 182 − 155 = 27

Box Q range = 190 − 140 = 50

Range = Highest class boundary − Lowest class boundary

e.g.

| Weight (kg) | 46-50 | 51-55 | 56-60 | 61-65 |
|---|---|---|---|---|
| Frequency | 4 | 7 | 11 | 8 |

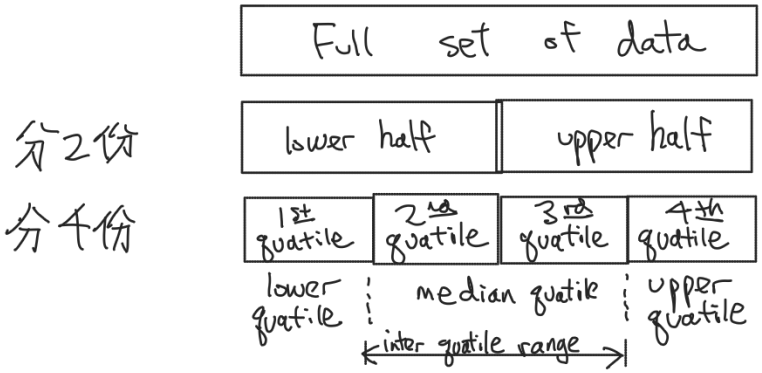
45.5 50.5 55.5 60.5 65.5

range = (65.5 − 45.5) kg = 20 kg

# Inter Quartile Range

↑

分四份




60.5 − 50.5 = 10.0

inter quatile range

e.g.

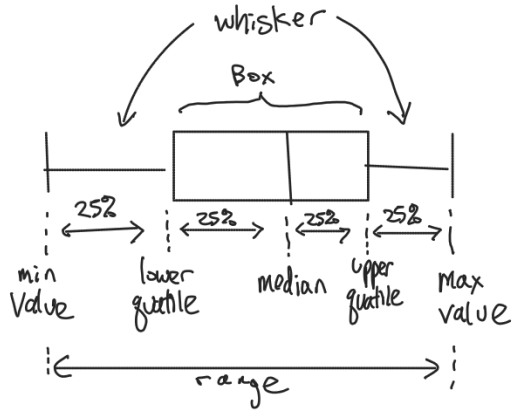
| Weight (kg) | 46-50 | 51-55 | 56-60 | 61-65 |
|---|---|---|---|---|
| Frequency | 4 | 7 | 11 | 8 |

45.5 50.5 55.5 60.5 65.5

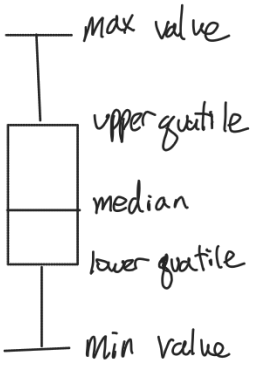
range = 65.5 − 45.5 = 20

↳ median = 55.5

# Box and whisker diagram




can also been drown as ...




# Standard Deviation ($\sigma$)

For ungroup data

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots (x_n - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$$

$\bar{x}$ : mean

$n$ : total number of data

$x_i$ : the $n^{th}$ data

$\sum$ : summation

variance

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

variance is a measure of dispersion

* in exam condition you are allowed to use calculator with statistics functions

* make sure you full master the skill of operating the calculator

# Standard Deviation for Grouped Data

$$\text{mean } \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{f_1 x_1 + f_2 x_2 \text{ ----- } f_n x_2}{f_1 + f_2 \text{ --- } f_n}$$

$f_i$ and $x_i$ are frequency and data of the ith class (from 1 to n)

example

| Scores | 11-20 | 21-30 | 31-40 | 41-50 |
|---|---|---|---|---|
| Class Mark | 15.5 | 25.5 | 35.5 | 45.5 |
| Frequency | 2 | 8 | 6 | 4 |

$$\text{mean } \bar{x} = \frac{(15.5 \times 2) + (25.5 \times 8) + (35.5 \times 6) + (45.5 \times 4)}{2 + 8 + 6 + 4} = 31.5$$

$$\sigma = \sqrt{\frac{2 \times (15.5 - 31.5)^2 + 8 \times (25.5 - 31.5)^2 + 6 \times (35.5 - 31.5)^2 + 4 \times (45.5 - 31.5)^2}{2 + 8 + 6 + 4}}$$

$$= \sqrt{84} = 9.17$$

3 significant digits

Comparison of dispersion measurement

| | | |
|---|---|---|
| range | only 2 data involved<br>easiest | may be misleading |
| interquatile range | only focus on middle 50%<br>no influence on extreme values | cannot show whole dispersion |
| standard deviation | takes all data into account<br>can show dispersion of whole data group | complex calculation involved |

# Applications of Standard Deviation

**standard score ($z$)** $\quad z = \dfrac{x - \bar{x}}{\sigma}$

a positive $z$ $\Rightarrow$ the value ($x$) is $z$ times standard deviation above the means

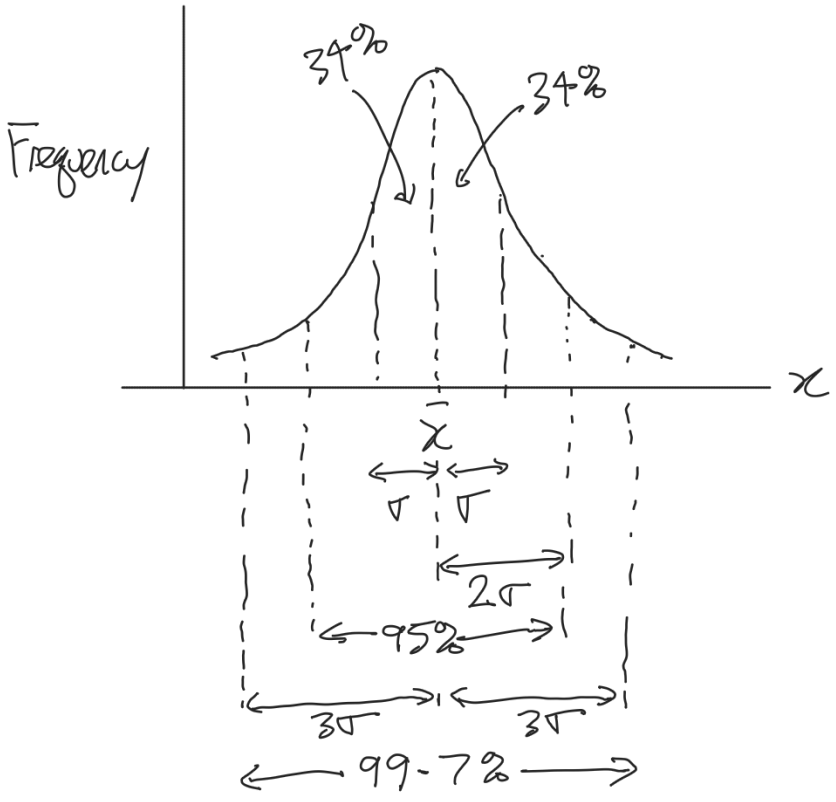
a negative $z$ $\Rightarrow$ the value ($x$) is $z$ times standard deviation below the means

e.g.: In a test $\bar{x} = 65$ $\sigma = 8$, John score 75

$$z = \frac{75 - 65}{8} = 1.25 \quad (1.25 \text{ times } \sigma \text{ above } \bar{x})$$

## Normal Distribution

Frequency distribution curve in most daily statistics is normal distribution




For normal distribution

mean = median = mode

# Effects on the dispersion with a change in data

* If a constant $k$ is added to every data in the set, then
  The range, interquartile range, and $\sigma$ will not change

* If $k$ is multiplied to every data in the set, then ....
  The range, interquartile range, and $\sigma$ will be multiplied by $k$

| | Range | Inter-Quad | Stand Dev |
|---|---|---|---|
| Add $k$ to each data | remain unchanged | | |
| Multiply $k$ to each data | becomes $k$ times the original | | |
| Remove a datum | decrease or unchange | may increase, decrease or remain unchange | |
| Insert a datum | increase or unchange | | |

$k$: constant