Representation of Data

Mean, Mode, Median

Mode is the value occurring the most often

Median is the middle value when all the data is arranged in order of size

Mean is found by adding together all the values of the data and dividing that total by the number of data values

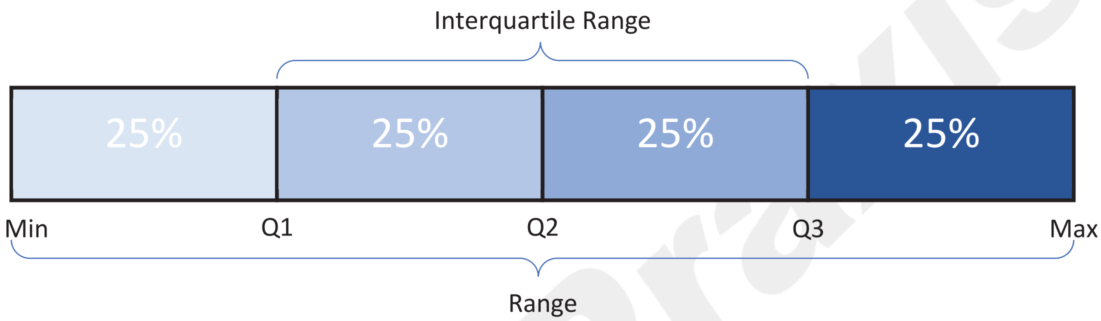
$$Ungrouped\ Data: \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$Grouped\ Data: \bar{x} = \frac{\sum_{i=1}^{n} x_i f}{\sum f}$$

Quartiles & Range

Range is the difference between the highest value and the lowest value in the data

Interquartile range is the difference between the upper quartile and lower quartile in the data



Standard Deviation & Variance

Variance

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{1}{n}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) = \frac{\sum x^2}{n} - (\bar{x})^2$$

Standard Deviation

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

Example:

The lengths (l cm) of a sample of nine otters, measured to the nearest cm by a wildlife research team, are:

76 77 91 65 63 83 92 61 88

Calculate the mean standard deviation of the nine recorded lengths

$$\bar{l} = \frac{\sum l}{n} = \frac{696}{9} = 77.3\ cm$$

$$\sigma = \sqrt{\frac{\sum l^2}{n} - (\bar{l})^2} = \sqrt{\frac{54998}{9} - \left(\frac{696}{6}\right)^2} = 11.4\ cm$$

Coding

If a data, x is coded using the formula

$$y = ax + b$$

The mean of the coded data, $\bar{y}$ would be

$$\bar{y} = a\bar{x} + b$$

The standard deviation of the coded data σ_y would be

$$\sigma_y = |a|\sigma_x$$

If an assumed mean, a, has been subtracted from all data values, x, then the formulae for the mean and standard deviation for the coded data will be

$$\overline{(x-a)} = \frac{\sum(x-a)}{n}$$

$$\sigma_{x-a} = \sqrt{\frac{\sum(x-a)^2}{n} - (\overline{x-a})^2}$$

Example:

A coffee machine is set to dispense 150 ml of coffee per cup.

In a random sample of 20 cups of coffee $\Sigma(c - 150) = -16$, where c ml is the volume of coffee in a cup.

(a) Find the mean volume in a cup of coffee.

(b) Given that $\Sigma(c - 150)^2 = 112$, find the standard deviation of the sampled cups of coffee.

$$\overline{c-150} = \frac{\sum(c-150)}{n} = -\frac{16}{20} = -0.8$$

$$\bar{c} = -0.8 + 150 = 149.2$$

$$\sigma_{c-150} = \sqrt{\frac{\sum(c-150)^2}{n} - \left(\overline{c-150}\right)^2}$$

$$= \sqrt{\frac{112}{20} - (-0.8)^2}$$

$$= \sqrt{4.96}$$

$$= 2.23$$

Steam and Leaf Diagrams

Drug 1

12 31 24 18 21 34 40 19 23 17 16

Drug 2

24 18 29 27 32 36 34 31 28 31

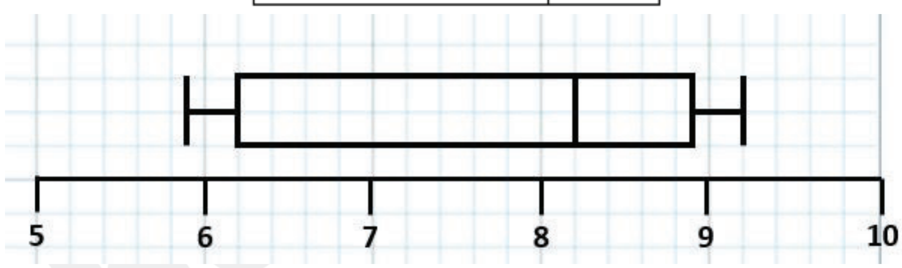
Final:

Drug 2		Drug 1
8	1	2 6 7 8 9
9 8 7 4	2	1 3 4
6 4 2 1 1	3	1 4
	4	0

Key: 8|1|2 means 18 for drug 2 and 12 for drug 1

Box Plots

Median	8.2
Lower Quartile	6.2
Upper Quartile	8.9
Lowest Value	5.9
Highest value	9.2



Cumulative Frequency

Type A		
Duration (h)	**Frequency**	**Cumulative frequency**
$0 \leq t < 5$	3	3
$5 \leq t < 10$	5	8
$10 \leq t < 15$	8	16
$15 \leq t < 20$	10	26
$20 \leq t < 25$	12	38
$25 \leq t < 30$	7	45
$30 \leq t < 35$	5	50

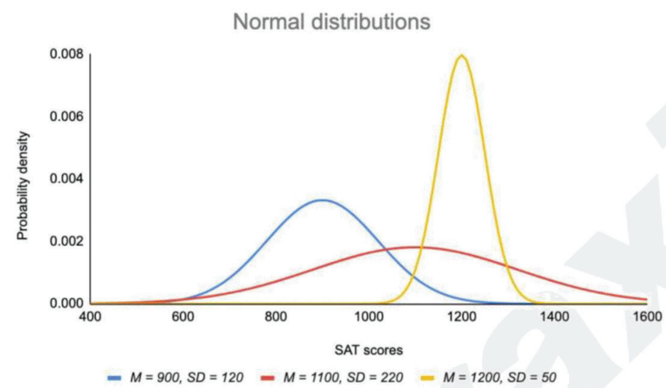
Normal Distribution

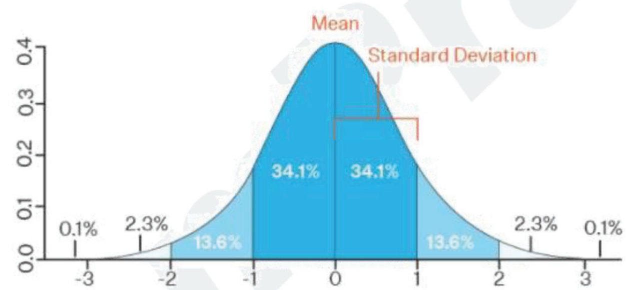
The continuous random variable can follow a normal distribution if:

- The distribution is symmetrical
- The distribution is bell-shaped

$$X \sim N(\mu, \sigma^2)$$

μ = mean
σ^2= variance





<u>Standard Normal Distribution</u>

The standard normal distribution is a normal distribution where the mean is 0 and the standard deviation is 1. It is denoted by

$$Z \sim N(0,1^2)$$

Where probability is

$$P(Z < z) = \phi(z)$$

Any normal distribution curve can be transformed to standard normal distribution curve using

$$Z = \frac{X - \mu}{\sigma}$$

Probability is related by

$$P(X < a) = P\left(Z < \frac{a - \mu}{\sigma}\right)$$

4 cases in terms of $\phi(z)$:

- $P(Z < z) = \phi(z)$
- $P(Z > z) = 1 - \phi(z)$
- $P(Z < -z) = 1 - \phi(z)$
- $P(Z > -z) = \phi(z)$