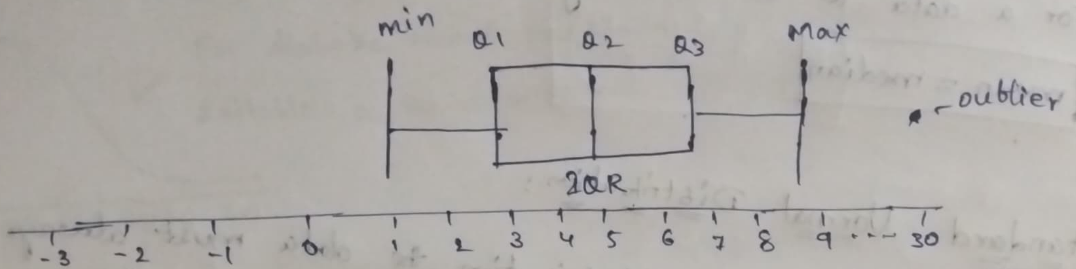
October-29:

IQR → Inter Quartile Range.

Box plot:



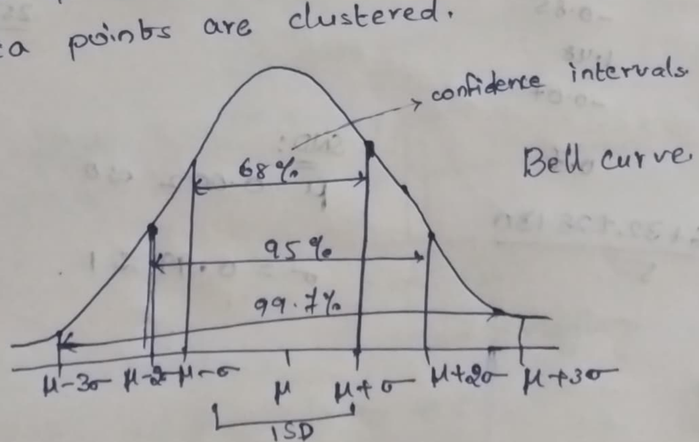
Any value below minimum & above maximum are called outliers.

Some ex's of box plot: The box plot doesn't look the same everytime, it varies based on distribution of data

Different Types of Distributions:

1. Normal Distribution (suitable for numerical data):

* The Normal distribution often called as "Guassian Distribution" is one of the most important probability distribution in Statistics & Machine Learning.
* It is "symmetric" & "Bell shaped", centered around the mean, where most of the data points are clustered.



* Emperical Rule:

1. 68% of the data lies within one standard deviation (+σ to -σ)
2. 95% of the data lies within two standard deviation (+2σ to -2σ)
3. 99.7% of the data lies within three standard deviation. (+3σ to -3σ)

Formula:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

probability density function. (PDF)

Example:
Human height, weight, test scores etc often follow a normal distribution.

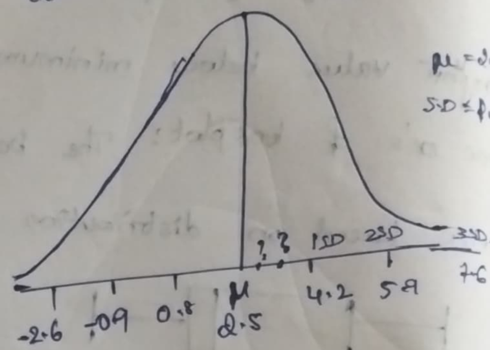
* For a data to be normally distribution it need to have its

$\boxed{\text{mean} = \text{median}}$

*

2. Standard Normal Distribution:

* in Standard Normal Distribution, the data must always possible



mean $\mu = 0$
standard deviation SD = 1.

* The SD varies based on your confidence interval.
for 68% → +1SD -1SD
for 50% → ? ... tough to calculate.

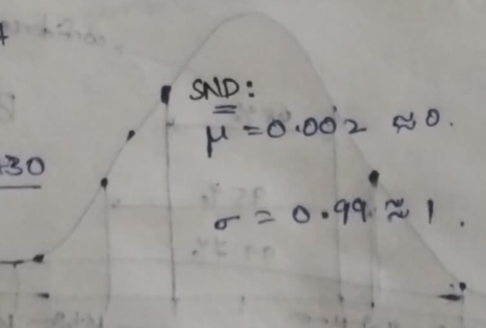
* To calculate the value for any point it is hard to go with normal distribution so we follow standard normal ~~devt~~ distribution

	ND Age	SND Age
Ex:	25	-1.25
	26	-0.85
	32	1.48
	28	-0.07
	30	0.7

$$\boxed{SND = \frac{x_i - \mu}{SD}}$$

$$\frac{25 - 28.2}{2.56}$$

$$\sum_{i=1}^{n}$$



SND:
$\mu = 0.002 \approx 0.$

$\sigma = 0.99 \approx 1.$

ND:
$$\mu = \frac{25+26+32+28+30}{5}$$

$$= \frac{141}{5}$$

$$\boxed{\mu = 28.2}$$

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

$$SD = \sqrt{V}$$
$$= \sqrt{\sigma^2}$$

$\sigma^2 = 6.56$

$\boxed{\sigma = 2.56}$