Calculation of sample mean, variance, st. dev.

Example-

Sample data on x = Temperature $n = 8$

x	x^2
-10	100
0	0
3	9
5	25
11	121
11	121
18	324
19	361
$\Sigma x = 57$	$\Sigma x^2 = 1061$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{57}{8} = 7.125^\circ$$

$$s^2 = \frac{(\Sigma x^2) - \frac{(\Sigma x)^2}{n}}{n-1} = \frac{1061 - \frac{(57)^2}{8}}{7}$$

$$s^2 = \frac{1061 - \frac{3249}{8}}{7} = \frac{1061 - 406.125}{7}$$

$$s^2 = \frac{654.875}{7} = 93.5536 \text{ sq}^\circ$$

$\bar{x} = 7.125^\circ$, $s^2 = 93.5536$ sq.°, $s = \sqrt{s^2} = \sqrt{93.5536}$
$= 9.672$

$s = 9.672^\circ$

Ex. 1.19 P. 44 Box Plot

X = AQI (Air Quality index)

2001 data from 15 cities ($n=15$) P. 15

Data set

X: 24, 3, 33, 14, 8, 31, 28, 4, 81, 19, 34, 50, 12, 6, 27

(1) List data values from L to H.

3, 4, 6, 8, 12, 14, 19, 24, 27, 28, 31, 33, 34, 50, 81

location for median $Q_2 = \frac{n+1}{2} = 8$ th (position data value)

$Q_2 = 24$

Lower half (below 8th position) has 7 data values

location for $Q_1 = \frac{7+1}{2} = 4$th data value

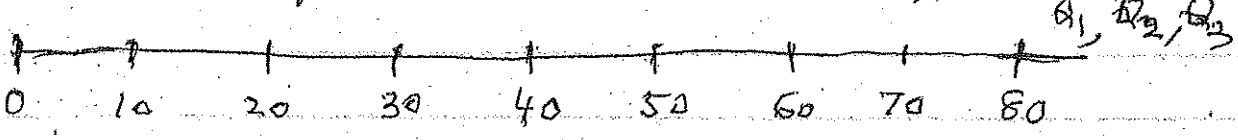
$Q_1 = 8$

Similarly upper half has also 7 data values and Q_3's location is 4th from top.

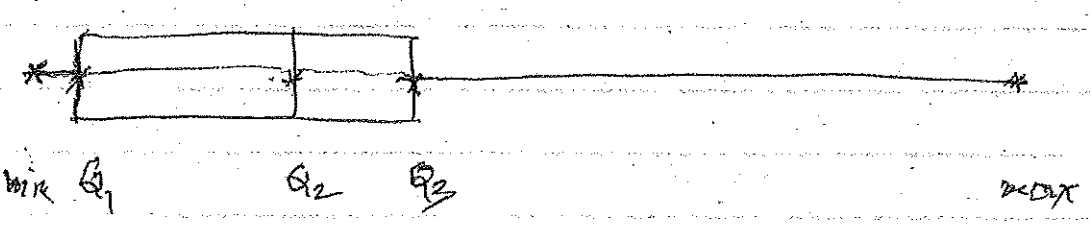
$Q_3 = 33$

$IQR = Q_3 - Q_1 = 33 - 8 = 25$

Scale for Box plot (5 numbers min = 3, max = 81)



Box Plot



Distn. is skewed to R

$1.5\,IQR = 1.5 * 25 = 37.5$

To determine outliers

$x > Q_3 + 1.5\,IQR$ i.e. $x > 33 + 37.5 = 70.5$ = upper bound

and

$x < Q_1 - 1.5\,IQR$ i.e. $x < 8 - 37.5 = -29.5$ lower bound

Since lower bound is negative we do not have any outliers on the lower side of data

On the upper side

max value $81 > 70.5 \rightarrow$ so 81 is an outlier

next higher value = 50 which is < 70.5

This is not an outlier.

There is only one outlier = 81