Calculation of sample means, variance, st. dev.

Example:

Sample data on \( x = \) Temperature, \( n = 8 \)

\[
\begin{array}{c|c|c|c}
\hline
x & x^2 & \bar{x} = \frac{\sum x}{n} = \frac{57}{8} = 7.125^\circ \\
-10 & 100 & \\
0 & 0 & \delta^2 = \frac{(\sum x^2) - \left(\frac{\sum x}{n}\right)^2}{n-1} = \frac{1061 - (57)^2}{8} \\
3 & 9 & \delta = \sqrt{\delta^2} = \sqrt{9.3336} = 9.672 \\
5 & 25 & \\
11 & 121 & \delta = \frac{1061 - 3249}{8} = \frac{1061 - 406.125}{7} \\
11 & 121 & \\
18 & 324 & \\
19 & 361 & \delta^2 = \frac{654.875}{7} = 93.5536^6 \bar{x} \\
\hline
\Sigma x = 57 & \Sigma x^2 = 1061 & \\
\hline
\end{array}
\]
Ex. 1.19 P. 44  Box Plot

X = AQI (Air Quality Index)

2001 data from 15 cities ($n=151$, P. 15
Data set

X: 24, 3, 33, 14, 8, 31, 28, 481, 19, 34, 59, 12, 5, 77

1) List data values from L to H.

3, 4, 6, 8, 12, 14, 19, 24, 27, 28, 31, 33, 34, 50, 581

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 = 581 $Q_1$, $Q_2$, $Q_3$)

Box Plot

$\min \leq Q_1 \leq Q_2 \leq Q_3 \leq \max$
Dist. is skewed to R
$1.5 \times IQR = 1.5 \times 25 = 37.5$

To determine outliers:
$x > Q_3 + 1.5IQR \iff x > 73 + 37.5 = 110.5$ upper bound

$x < Q_1 - 1.5IQR \iff x < 8 - 37.5 = -29.5$ lower bound

Since lower bound is negative we don't have any outliers on the lower side of data.

On the upper side:
max value 81 > 110.5 so 81 is an outlier.

next higher value = 50 which is < 70.5
This is not an outlier.

There is only one outlier = 81