

# Sampling

## Statistics

- A statistic is used to estimate a parameter.

using  $\sqrt{\frac{\sum (oc - \bar{oc})^2}{n}}$  to estimate the standard deviation

of the population will give a smaller spread (closer to the sample than the population) HENCE we use

$$s = \sqrt{\frac{\sum (oc - \bar{oc})^2}{n-1}}$$

to achieve a slightly larger no.

↑  
unbiased estimator of the population standard deviation

## Types of Sampling

- Simple random.
- Systematic.
- Stratified.
- Opportunity.
- Quota.
- Cluster.

## Quartiles & Outliers

- Median =  $\left(\frac{n+1}{2}\right)^{th}$  value (if in between then mean of values either side)

- Outliers are either  $< Q_1 - 1.5 IQR$  or  $> Q_3 + 1.5 IQR$

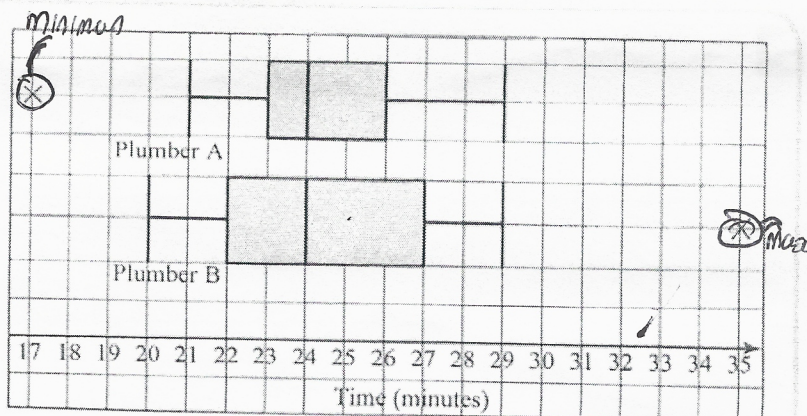
- $Q_1 = \frac{n+1}{4}^{th}$  value &  $Q_3 = \frac{3(n+1)}{4}^{th}$  value



A building company works with two plumbers. Over a period of time, they assess how long it takes each plumber to fix leaking pipes.

This data is displayed in box-and-whisker plots.

An outlier is defined as a value less than  $Q_1 - 1.5 \times \text{IQR}$  or more than  $Q_3 + 1.5 \times \text{IQR}$ .



- a Write down the minimum, lower quartile, median, upper quartile and maximum for each set of data.
- b Recommend a choice of plumber given that no outliers are deleted.

*Suggesting their best choice.*

a minimum /  $P_n = 21 \times 17$  LQ /  $P_n = 23$  Median /  $P_n = P_B = 24$   
 $P_B = 20$  ( $Q_1$ ) /  $P_B = 22$  ( $Q_2$ )

UQ /  $P_n = 26$  Maximum /  $P_n = 29$   
 $(Q_3) / P_B = 27$   $P_B = 29 \times 35$

b Plumber A  $\because$  they have a time of 17 mins as an outlier & a less variable IQR and the same median.

Table of  
+ & -

	Advantages	Disadvantages
Box and Whisker Plots	<ul style="list-style-type: none"> <li>Can see how variable the data is</li> <li>+ Easy to compare side by side.</li> <li>+ Symmetry</li> <li>+ Clear outliers, clustered groupings</li> </ul>	<ul style="list-style-type: none"> <li>Cannot see how dense the data is.</li> <li>- Emphasis on outliers</li> </ul>
Cumulative Frequency Diagrams	<ul style="list-style-type: none"> <li>+ Easy to find percentages</li> <li>+ Information is strong &amp; clear</li> <li>+ Shows running totals</li> </ul>	<ul style="list-style-type: none"> <li>- Hard to compare individual frequencies at a glance</li> </ul>