Pearson BTEC Levels 4 and 5 Higher Nationals in Engineering (RQF)

Unit 2: Engineering Maths (core)

Unit Workbook 2

in a series of 4 for this unit

Learning Outcome 2

Statistical Techniques

2.1 Tabular and Graphical Form

2.1.1 Data Collection Methods

Data can be collected in a number of ways, including …

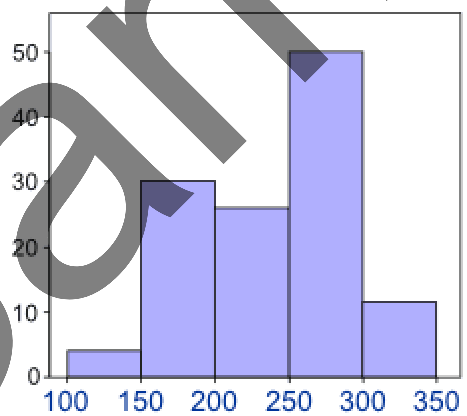
- Using test instruments such as multimeters, frequency counters, strain gauges etc.
- By observation – i.e. counting events, performance analysis
- Software simulation
- Wireless telemetry
- Artificial Intelligence
- Mobile phone applications

Continuous data can have any value, with potentially an unlimited number of decimal places in the data. An example is the voltage reading on an analogue voltmeter. *Discrete* data is obtained by some form of counting, such as recording the number of products on a production line.

There a several ways of presenting acquired data. These are discussed below.

2.1.2 Histograms

A histogram is a representation of data which is grouped into ranges. An example histogram is shown below…



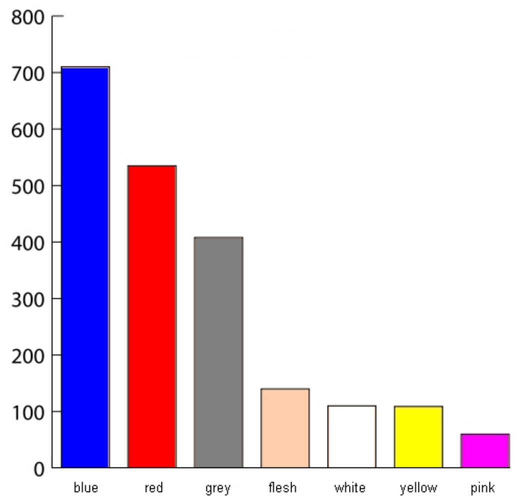
Here it can be seen, for example, that there were 30 occurrences of recorded data between 150 and 200 units.

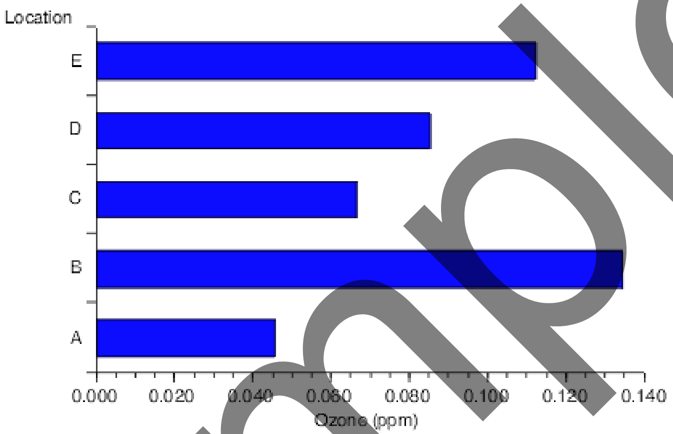
2.1.3 Bar Charts

A bar chart can represent data in a similar way to a histogram. There are two ***differences between a bar chart and a histogram***…

- A bar chart has spaces between the rectangles. A histogram does not.
- A bar chart can represent data as either horizontal or vertical rectangles. Histograms are vertical.

Examples of bar charts are shown below…



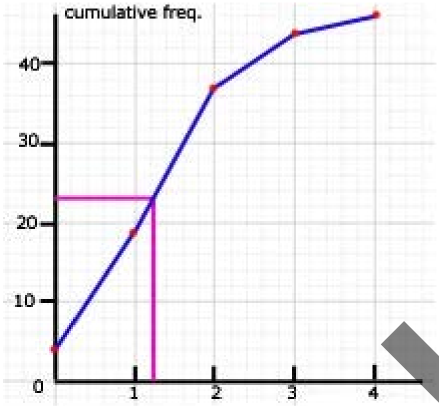


2.1.4 Line Diagrams

These represent discrete transitions to new data, marked by straight lines to specific data points, as per the example below…

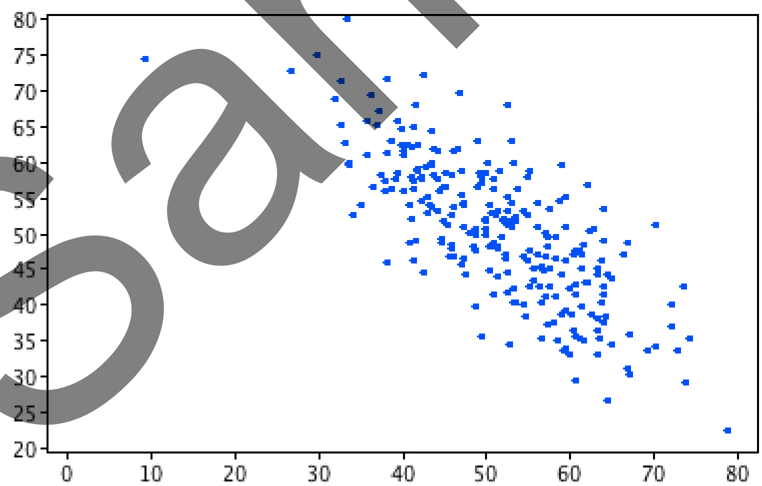
2.1.5 Cumulative Frequency Distribution Diagrams

These can be thought of as a running total in ascending order of the units on the horizontal axis. An example is shown below…



2.1.6 Scatter Plots

These are plotted with dots at matching vertices. Usually a line of 'best fit' may be drawn through the points. This line can be constructed using an equation or by estimation. A typical scatter plot is shown below…



2.1.7 Tally Charts

These are used to record observed data in groups. A vertical line is used to record an event. Usually a fifth event is recorded by inserting a diagonal line through the previous four vertical lines, as per the example below…

Mark	*Tally*	*Frequency*
2	|	1
3	||	2
4	|||	3
5	|||	3
6	~~||||~~	5
7	~~||||~~ |	6
8	~~||||~~ ||	7
9	||	2
10	|	1

Worked Example 1

The age of ten random students was recorded. The data is as follows…

Sample	1	2	3	4	5	6	7	8	9	10
Age (Years)	**19**	**18**	**20**	**19**	**18**	**19**	**19**	**23**	**19**	**19**

a) Produce a Tally Chart to show the frequency of the students' ages.
b) Produce a Cumulative Frequency Distribution Curve for this data set.

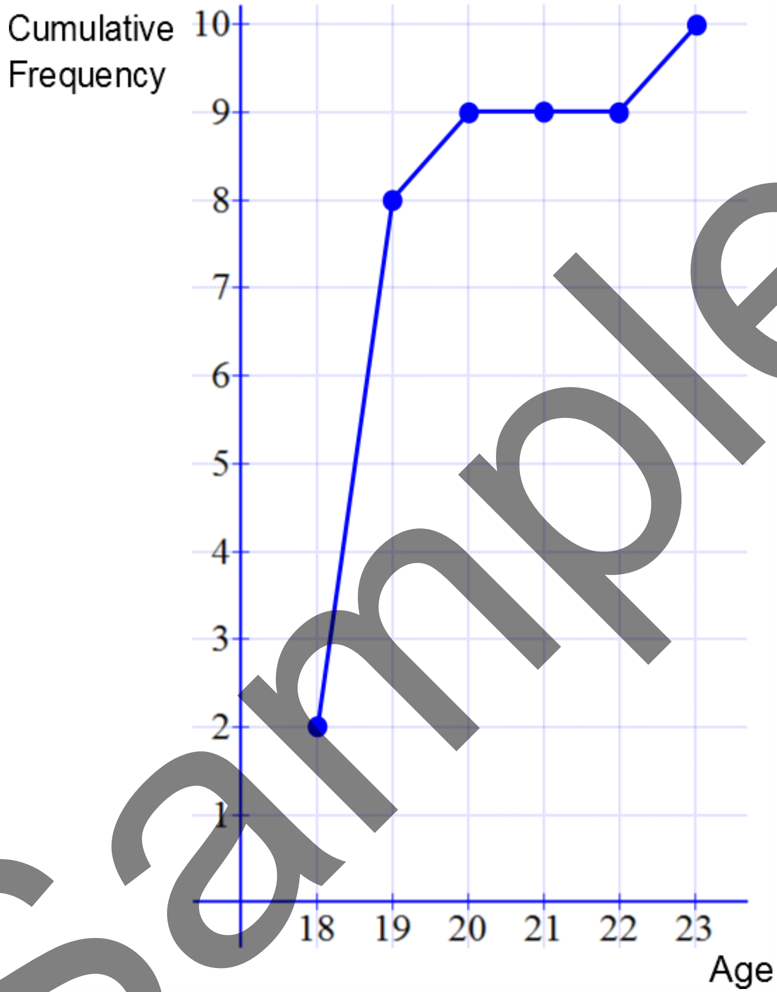
(a)

We see from the data that the youngest observed student is 18 years old and the oldest is 23 years old. These numbers are the bounds of our data in the tally chart…

Age	Tally	Frequency
18	||	2
19	~~||||~~ |	6
20	|	1
21		0
22		0
23	|	1

(b)

We may start our axes anywhere we like. For this plot we choose the vertical and horizontal axes as shown below. Notice that the 'curve' never assumes a negative slope at any point since it is cumulative (a running total).



Question

Q2.1 **The resistance of 10 random samples from a 22Ω resistor production line was recorded. The data is as follows…**

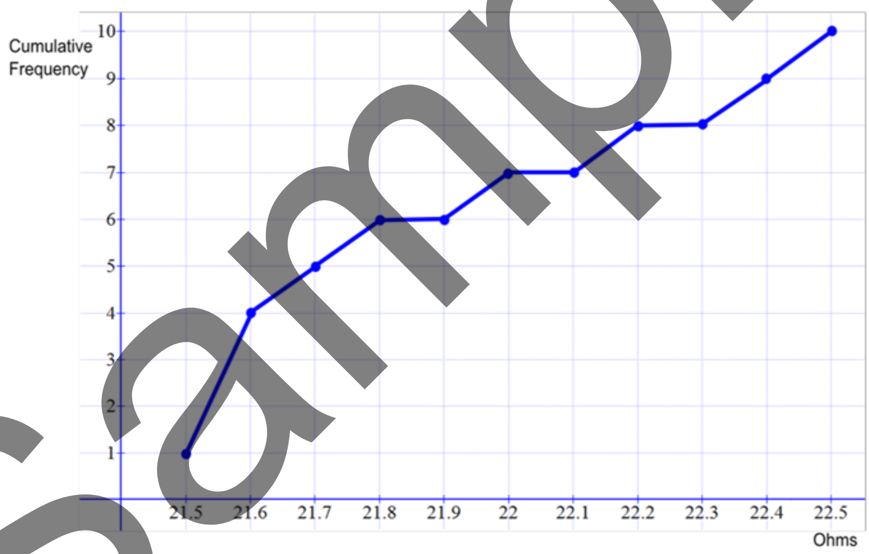
Sample	1	2	3	4	5	6	7	8	9	10
Resistance (Ω)	**21.5**	**22.5**	**21.6**	**21.6**	**21.6**	**22.4**	**22.0**	**22.2**	**21.7**	**21.8**

a) Produce a Tally Chart to show the frequency of the resistances.

b) Produce a Cumulative Frequency Distribution Curve for this data set.

ANSWER

Resistance (Ω)	Tally	Frequency
21.5	|	1
21.6	|||	3
21.7	|	1
21.8	|	1
21.9		
22.0	|	1
22.1		
22.2	|	1
22.3		
22.4	|	1
22.5	|	1



2.2 Central Tendency and Dispersion

2.2.1 The Concept of Central Tendency and Variance Measurement

A set of data may be represented by a single value. You will have most commonly used the term 'average' for a representative size a set of data. The correct mathematical term for this is the '*arithmetic mean*' or just '*mean*'.

On some occasions it might be more useful to represent the *most common* term in a set of data. This is termed the '*mode'* of the data. Another way is to order the data in ascending order and use the middle value, the *median*, as a representation of the data. A representative method of finding the central tendency of a data set is chosen to best reflect the nature of the data.

The idea of the *variance* of a set of data aims to quantify how much the data varies from the central point.

2.2.2 Mean, Median and Mode

Let us again consider the data from worked example 1 (reproduced below) to find the mean, median and mode…

Sample	1	2	3	4	5	6	7	8	9	10
Age (Years)	**19**	**18**	**20**	**19**	**18**	**19**	**19**	**23**	**19**	**19**

The ***mean*** (average) is simply calculated by summing all the data and dividing by the number of data samples. Expressed mathematically, the mean is…

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{10} x_i = \frac{19+18+20+19+18+19+19+23+19+19}{10} = 19.3$$

Notice that we use a bar over the x to represent its mean.

The ***median*** is calculated by ordering the data in ascending order and using the middle value, the median, as a representation of the data…

18, 18, 19, 19, ***19, 19***, 19, 19, 20, 23

Since we do not have an odd number of samples here then we cannot directly select the central value and call that the median. In situations like this we simply select the two central values (bold) of the ascending list and average them to find the median…

$$median = \frac{19+19}{2} = 19$$

To find the ***mode*** of the data we simply look for the most common data item. Here, this is clearly 19. Sometimes data will have two or more most common values, in which case the data is termed *bimodal* or *multimodal*.

2.2.3 Standard Deviation and Variance

To ascertain the amount of variation in a full given population, or sample of a population, we use the terms *standard deviation* and *variance*. Let's define these terms…

$$\boldsymbol{standard\ deviation\ for\ a\ {\color{red}full}\ population, \sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}}$$

$$\boldsymbol{standard\ deviation\ {\color{red}sample}\ of\ population, \sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}}$$

$$\boldsymbol{variance = \sigma^2}$$

The Greek lower case letter sigma (σ) is used for standard deviation and n is the number of samples. You will notice that the variance is simply the square of the standard deviation. Let's see these in use.

Worked Example 2

Find the standard deviation and variance for the following population data set…

Sample	1	2	3	4	5	6	7	8	9	10
Age (Years)	**19**	**18**	**20**	**19**	**18**	**19**	**19**	**23**	**19**	**19**

We shall firstly find the standard deviation and then square that answer to get the variance. To find the standard deviation we normally construct a table, as follows…

$\boldsymbol{x_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{(x_i - \bar{x})^2}$
19	19.3-19=0.3	0.09
18	19.3-18=1.3	1.69
20	19.3-20=-0.7	0.49
19	19.3-19=0.3	0.09
18	19.3-18=1.3	1.69
19	19.3-19=0.3	0.09
19	19.3-19=0.3	0.09
23	19.3-23=-3.7	13.69
19	19.3-19=0.3	0.09
19	19.3-19=0.3	0.09
$\bar{x} = \frac{1}{n}\sum_{i=1}^{10} x_i = 19.3$		$\sum(x_i - \bar{x})^2 = 18.1$

$$\therefore \quad \boldsymbol{standard\ deviation, \sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{18.1}{10}} = \sqrt{1.81} = 1.345}$$

$$\therefore \quad \boldsymbol{variance = \sigma^2 = 1.81}$$

Question

Q2.2 (a) Calculate the mean, median and mode for the data set below.

(b) Find the standard deviation and variance for the population data set.

Sample	1	2	3	4	5	6	7	8	9	10
Salary (£1000's)	**22**	**24**	**21**	**26**	**32**	**26**	**26**	**54**	**12**	**33**

ANSWER

Q2.2

(a) Mean = £27,600, Median = £26,000, Mode = £26,000
(b) Standard Deviation = 10.41, Variance = 108.44

Check your answers with this handy online calculator.

2.2.4 Interquartile Range

Dispersion is sometimes measured in quartile values. Consider the data set…

1, 2, 2, 4, 6, 8, 9, 10, 10, 11, 12, 12, 13, 14, 17

This may be broken up into four quarters. To do this we need to make three splits. The points of these splits are known as the quartile values Q_1, Q_2 and Q_3. The value of Q_2 is the median and the other quartiles are determined by finding the medians to the left and right of Q_2. The data set below has these three quartile values highlighted in red…

1, 2, 2, **4**, 6, 8, 9, **10**, 10, 11, 12, **12**, 13, 14, 17

So, $Q_1 = 4, Q_2 = 10$ and $Q_3 = 12$. There is a range associated with a pair of these quartiles, known as the *interquartile range*. In this example the first interquartile range is (10 -4) = 6 and the second interquartile range is (12 – 10) = 2.

2.2.5 Application to Engineering Production

Much of the material covered so far in this workbook is used very frequently in quality control departments of manufacturing and assembly plants.

Very often in industry there will be grouped samples of data taken and these samples could be taken at random times or regular times. Each company will have its own in-house method of quality control and also data presentation.

Note that the formula for standard deviation of an incomplete sample is different than that for a full population. This is explained in the handy online calculator.