**Statistics for Science 1**

**Outcome 1**

**Analysing and Grouping Data**

In this outcome, we shall be looking at ways of representing data in various forms by a single number (average) and also giving an indication of how the data is spread out.

We shall also investigate various ways of representing data in a graphical manner depending on how the data is organised.

**1. Raw Data**

This is the data represented as a list of numbers. This would be how the data would look after it had been collected.

The Range

This is an indication of how the data is spread out. It is simply the biggest number - the smallest number.

e.g. The following data relates to the hours of sunshine in a holiday resort over one week :

Days Mon Tue Wed Thu Fri Sat Sun

Hours of 7 5 3.5 5.75 7 8 9

Sunshine

 The range = 9 - 3.5 = 5.5 hours

Population and Sample

A population is all the possible data that could be collected.

A sample is some of the population that has been chosen to represent the population. There are many ways to collect a sample but we shall not be studying these in this unit.

There is a difference in the symbols of the mean and the formula for the standard deviation whether the data is a sample or a population.

This will always be clearly stated in the question.

Mean

To find the mean from raw data, we add up all the numbers and divide the total bu the number of numbers.

If we call the numbers x, then adding up all the numbers is shown as ∑x [the sum of x]. The number of numbers is called n. The mean for a sample is $\overbar{x}$. The mean for a population is µ.

We now have the formulae :

 $\overbar{x}= \frac{\sum\_{}^{}x}{n}$ $μ= \frac{\sum\_{}^{}x}{n}$

Note that there is no actual difference in the mean whether it is a sample or a population.

Standard Deviation

This is a way of showing how the data is arranged around the mean. We shall study how this is used in more detail in Outcome 2.

The idea is to find how each number deviates from the mean and average out this deviation.

The standard deviation for a sample is called s and the standard deviation for a population is called σ.

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

$$σ= \sqrt{\frac{\sum\_{}^{}x^{2 }- \frac{(\sum\_{}^{}x)^{2}}{n}}{n}}$$

The Coefficient of Variation The formula is V = $[\frac{σ}{μ} ×100]$ %

Note that the units for means and standard deviations are as the question but the coefficient of variation is a percentage. It shows the extent of variability in relation to the mean of the population.

Standard Error The formula is S.E = $\frac{σ}{\sqrt{n}}$

Example : Find the mean and standard deviation for the following data as a sample and as a population and the coefficient of variation and standard error for the population:

 The following data relates to the hours of sunshine in a holiday resort over one week :

 Days Mon Tue Wed Thu Fri Sat Sun

 Hours of 7 5 3.5 5.75 7 8 9

 Sunshine

Sample : ∑x = 45.25 n = 7 $\overbar{x}$ = $\frac{45.25}{7}$ = 6.46 hours

 ∑x2 = 313.3125 s = $\sqrt{\frac{313.3125- \frac{45.25^{2}}{7}}{6}}$ = 1.86 hours

Population : ∑x = 45.25 n = 7 µ= $\frac{45.25}{7}$ = 6.46 hours

 ∑x2 = 313.3125 σ = $\sqrt{\frac{313.3125- \frac{45.25^{2}}{7}}{7}}$ = 1.72 hours

V = $\frac{1.72}{6.46}$ x 100 %= 26.63 %

S.E = $\frac{1.72 }{\sqrt{7}}$ = 0.65

Exercise 1 :

For the following questions, find (a) the range, (b) the mean and standard deviation as a sample, (c) the mean and standard deviation as a population (d) the coefficient of variation and standard error for the population:

1. The following data shows the reaction times (in ms) of 8 people:

 1, 3, 5, 5, 6, 7, 7, 9

2. 2.3 3.7 2.9 3.1 2.8 2.7 3.0

3. 5.3 6.2 4.9 6.3 6.2 5.8 5.7 (in m)

4. 352 279 302 285 298 301 (in kg)

We are now going to look at another common method of analysing data and the graph that is connected with this analysis.

**Median and Quartiles**

Because the mean has to include every number it can be skewed by a very small or a very big piece of data so we need another way of finding an average.

If a set of **n** values are arranged in ascending or descending order, then the **median** is the middle value (if **n** is odd) or the mean of the two middle numbers (if **n** is even).

So if given 11 items arranged in order, the median is the 6th item.

If given 10 items the median is the mean of the 5th and 6th.

The **position of the median** can be calculated by: Median Position = 

A measure of how the data is arranged is found by finding the quartiles - these are the Lower Quartile, Q1, which is the number one quarter of the way along and the Upper Quartile, Q3, which is the number three quarters of the way along.

Example 1 : Find the median and quartiles of 8, 6, 4, 2, 2, 5, 8

 Answer : Rearrange : 2, 2, 4, 5, 6, 8, 8

 median = 5 Q1 = 2 Q3 = 8

Example 2 : Find the median and quartiles of 8, 6, 4, 3, 2, 6, 7, 8

 Rearrange : 2, 3, 4, 6, 6, 7, 8, 8

 median = 6 (lies between the two ‘middle sixes’)

 Q1 = $\frac{3+4}{2}= $3.5 Q3 = $\frac{7+8}{2}$ = 7.5

Example 3 : Find the median and quartiles of 8, 20, 16, 19, 23, 12, 14, 15

 Rearrange : 8, 10, 12, 14, 15, 16, 19, 23

 median = $\frac{14+15}{2}$ = 14.5 (14.5 is the mean of the two middle numbers)

 Q1 = $\frac{10+12}{2}$ = 11 Q3 = $\frac{16+19}{2}$ = 17.5

**Boxplots**

This is a graph that is used specifically with the median and quartiles.

Box plots are used to illustrate a distribution using the maximum, minimum, median and quartile values. Box plots can be used to help compare distributions.

Min Q1 Q2 Q3 Max

Example

 Min = 16, Max = 94 Q1 = 36 Q2 = 48 Q3 = 60

 for an exam which was out of a100. The box plot is:

 10 20 30 40 50 60 70 80 90 100

Note that these should be drawn to scale on graph paper.

## Exercise 2

1. For each data set, write down the minimum, maximum, median, upper and lower quartiles and draw a box plot.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a. | 19 | 27 | 12 | 30 | 8 | 31 | 25 |  |  |  |  |  |  |
| b. | 4 | 7 | 10 | 2 | 6 | 4 | 14 | 8 | 15 |  |  |  |  |
| c. | 4.0 | 2.9 | 5.3 | 1.8 | 4.0 | 4.7 | 2.8 | 1.8 | 5.2 | 4.0 | 5.1 |  |  |
| d. | 18 | 11 | 12 | 11 | 16 | 20 | 10 | 15 | 13 | 14 | 15 |  |  |
| e. | 51 | 58 | 53 | 51 | 52 | 55 | 53 | 50 | 54 | 53 | 52 |  |  |
| f. | 249 | 265 | 254 | 267 | 270 | 279 | 252 | 268 | 258 |  |  |  |  |
| g. | 82 | 90 | 97 | 85 | 105 | 86 | 96 | 104 | 108 | 94 | 96 |  |  |
| h. | 40 | 43 | 41 | 41 | 40 | 50 | 40 | 44 | 80 | 40 | 41 | 40 |  |
| i. | 0.1 | 0.8 | 0.3 | 0.2 | 0.2 | 0.5 | 0.3 | 0.1 | 0.4 | 0.3 | 0.2 |  |  |
| j. | 29 | 25 | 13 | 39 | 29 | 26 | 18 | 18 | 33 | 31 | 19 | 30 | 26 |

2. Here are two sets of marks for a French test.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 98 | 94 | 92 | 78 | 88 | 78 | 82 | 98 | 68 | 66 |
| 100 | 96 | 84 | 86 | 84 | 94 | 86 | 92 | 82 | 100 |

**Class 5A**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 73 | 95 | 80 | 72 | 85 | 90 | 91 | 88 | 91 | 93 |
| 83 | 76 | 93 | 75 | 88 | 94 | 88 | 91 | 91 | 75 |

**Class 5B**

Draw a box plot for each class and compare the results.

3. A company that manufactures shoelaces spot checks the length (in cm) of the laces.

 Here are the results for two different production lines.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Line A | 26.8 | 27.2 | 26.5 | 27.0 | 27.3 | 27.5 | 26.1 | 26.4 | 27.9 | 27.3 |
| Line B | 26.8 | 26.7 | 27.1 | 27.0 | 26.9 | 27.0 | 27.3 | 26.9 | 27.0 | 27.3 |

 Draw a box plot for line A and line B.

 Which is the better production line ? (Give a reason for your answer)

4. Two sixth year classes take part in a Sponsored Fast for Famine Relief. The number of hours each pupil lasted are shown below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 6C1 | 20 | 22 | 21 | 20 | 22 | 20 | 22 | 20 | 20 | 24 | 21 | 22 | 23 | 22 | 22 | 23 |
| 6C2 | 15 | 20 | 24 | 23 | 22 | 24 | 18 | 24 | 22 | 23 | 24 | 17 | 20 | 24 | 24 | 20 |

 Show each class on a box plot and comment on any differences.

**Answers**

Exercise 1

1. (a) range = 9 - 1 = 8 ms

 (b) $\overbar{x}$ = $\frac{43}{8}$ = 5.375 ms

 s = $\sqrt{\frac{275- \frac{43^{2}}{8}}{7}}$ = 2.50 ms

 (c) µ = 5.375 ms

 σ = $\sqrt{\frac{275- \frac{43^{2}}{8}}{8}}$ = 2.34 ms

 (d) V = $\frac{2.34}{5.375}$ x 100 = 43.53 % S.E =$\frac{2.34}{\sqrt{8}}$ = 0.83

2. (a) range = 3.7 – 2.3 = 1.4

 (b) $\overbar{x}$ = $\frac{20.5}{7}$ = 2.93

 s = $\sqrt{\frac{61.13- \frac{20.5^{2}}{7}}{6}}$ = 0.43

 (c) µ = 2.93

 σ = $\sqrt{\frac{61.13- \frac{20.5^{2}}{7}}{7}}$ = 0.40

 (d) V = $\frac{0.4}{2.93}$ x 100 % = 13.65 % S.E =$\frac{0.4}{\sqrt{7}}$ = 0.15

3. (a) range = 6.3 – 4.9 = 1.4 m

 (b) $\overbar{x}$ = $\frac{40.4}{7}$ = 5.77 m

 s = $\sqrt{\frac{234.8- \frac{40.4^{2}}{7}}{6}}$ = 0.52 m

 (c) µ = 5.77 m

 σ = $\sqrt{\frac{234.8- \frac{40.4^{2}}{7}}{7}}$ = 0.48 m

 (d) V = $\frac{0.48}{5.77}$ x 100 = 8.32 % S.E =$\frac{0.48}{\sqrt{7}}$ = 0.18

4. (a) range = 352 – 279 = 73 kg

 (b) $\overbar{x}$ = $\frac{1817}{6}$ = 302.8 kg

 s = $\sqrt{\frac{553579- \frac{1817^{2}}{6}}{5}}$ = 25.81 kg

 (c) µ = 302.8 kg

 σ = $\sqrt{\frac{553579- \frac{1817^{2}}{6}}{6}}$ = 23.56 kg

 (d) V = $\frac{23.56}{302.8}$ x 100 = 7.78 % S.E =$\frac{23.56}{\sqrt{6}}$ = 9.61

Exercise 2

1. (a) 8 12 19 25 27 30 31

 min : 8 median : 25

 max : 31 Q1 : 12 Q3 : 30

 **a.**

 

 (b) 2 4 4 6 7 8 10 14 15

 min : 2 median : 7

 max : 15 Q1 : 4 Q3 : 12

 

 (c) 1.8 1.8 2.8 2.9 4.0 4.0 4.0 4.7 5.1 5.2 5.3

 min : 1.8 median : 4.0

 max : 5.3 Q1 : 2.8 Q3 : 5.1



 (d) 10 11 11 12 13 14 15 15 16 18 20

 min : 10 median : 14

 max : 20 Q1 : 11 Q3 : 16

 

 (e) 50 51 51 52 52 53 53 53 54 55 58

 min : 50 median : 53

 max : 58 Q1 : 51 Q3 : 54

 

 (f) 249 252 254 258 265 267 268 270 279

 min : 249 median : 265

 max : 279 Q1 : 253 Q3 : 269

 

 (g) 82 85 86 90 94 96 96 97

 104 105 108

 min : 82 median : 96

 max : 108 Q1 : 86 Q3 : 104

 

 (h) 40 40 40 40 40 41 41 41 43

 44 50 80

 min : 40 median : 41

 max : 80 Q1 : 40 Q3 : 43.5

 

 (i) 0.1 0.1 0.2 0.2 0.2 0.3 0.3 0.3 0.4 0.5 0.8

 min : 0.1 median : 0.3

 max : 0.8 Q1 : 0.2 Q3 : 0.4

 

 (j) 13 18 18 19 25 26 26 29 29

 30 31 33 39

 min : 13 median : 26

 max : 39 Q1 : 18.5 Q3 : 30.5



2, (a) 66 68 78 78 82 82 84 84 86 86

 88 92 92 94 94 96 98 98 100 100

 min : 66 median : 87

 max : 100 Q1 : 82 Q3 : 95

 (b) 72 73 75 75 76 80 83 85 88 88

 88 90 91 91 91 91 93 93 94 95

 min : 72 median : 88

 max : 95 Q1 : 78 Q3 : 91



3. (A) 26.1 26.4 26.5 26.8 27.0 27.2 27.3 27.3 27.5 27.9

 min : 26.1 median : 27.1

 max : 27.9 Q1 : 26.5 Q3 : 27.3

 (B) 26.7 26.8 26.9 26.9 27.0 27.0 27.0 27.1 27.3 27.3

 min : 26.7 median : 27.0

 max : 27.3 Q1 : 26.9 Q3 : 27.1



4. 6C1

 20 20 20 20 20 21 21 22 22 22 22 22 22 23 23 24

 min : 20 median : 22

 max : 24 Q1 : 20 Q3 : 22

 6C2

 15 17 18 20 20 20 22 22 23 23 24 24 24 24 24 24

 min : 15 median : 22.5

 max : 24 Q1 : 20 Q3 : 22.5

