# Resource for Day 10 – Data Handling – Wednesday 1 November

1. When a rugby player is substituted the average weight of the 8 players in the scrum drops by 2 kg. How much less did he weigh than the player he replaced?

2. Find, to 3 d.p. the standard deviation of the following numbers without using a calculator showing all working details:

   2 ; 5 ; 8 ; 13

3. Thabo has an average of 78% for his 5 maths tests so far this term. What mark must he get on the 6$^{th}$ and final test for his average mark to be an A (80%)?

4. Consider the data represented in this stem and leaf diagram:

| | |
|---|---|
| 3 | 3 8 |
| 4 | 0 3 3 3 3 5 6 |
| 5 | 0 2 3 3 5 5 5 6 7 8 9 9 |
| 6 | 1 4 4 7 8 |
| 7 | 6 |

   a. Determine the range

   b. Detemine the mean to 2 d.p.

   c. Determine the standard deviation to 2 d.p.

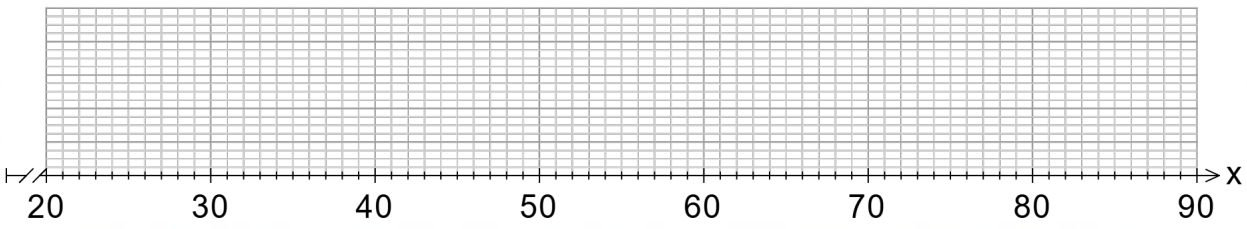
   d. Determine the mode

   e. Determine the 5 number summary

   f. Determine the inter-quartile range

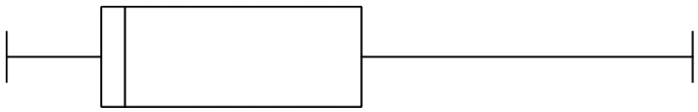
   g. If an outlier is deemed to be a data value which is more than 1,5 times the inter-quartile range above $Q_3$ or more than 1,5 times the inter-quartile range below $Q_1$ then determine whether the above data set has any outliers?

   h. Draw a box and whisker plot on the axes provided



5. Describe the symmetry or skewness of each of the following data sets:

a.



b.

c.

6. Consider the grouped data below

| Group | frequency |
|---|---|
| 35-45 | 1 |
| 45-55 | 3 |
| 55-65 | 4 |
| 65-75 | 14 |
| 75-85 | 10 |
| 85-95 | 13 |
| 95-105 | 12 |
| 105-115 | 13 |
| 115-125 | 7 |
| 125-135 | 3 |

a. Determine the modal group.

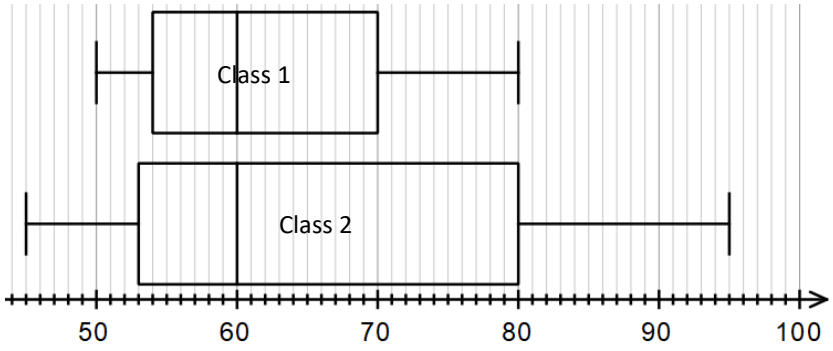
b. In which group will the median lie?

c. In which group will the first quartile lie?

d. In which group will the third quartile lie?

e. Determine, to 1 d.p. an estimate for the mean and explain why it is only an estimate.

7. The marks on a Maths test for two different classes are displayed in the box and whisker plots below:




a. Determine the median mark for class 1.

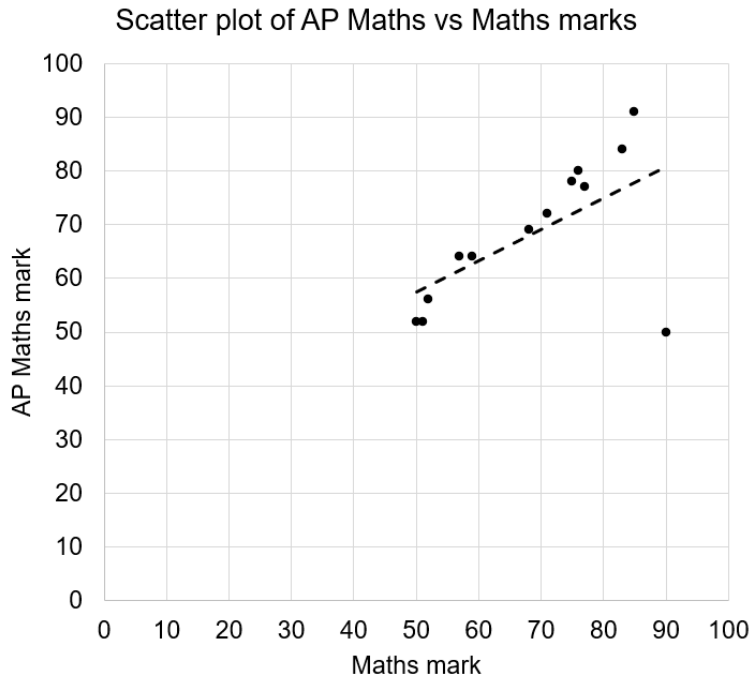
b. Determine the inter-quartile range for class 2.

c. Describe the skewness / symmetry of the data in both classes, justifying your answer.

d. Give $p$ if 75% of the candidates in class 1 achieved a mark of greater than $p$.

e. Which class has the greater standard deviation? Justify your answer.

8. The Maths and AP Maths marks of 13 students are plotted below:




a. Comment on the nature and strength of the relationship between the performance in the two subjects.

b. The teacher plans to use the equation of the least squares regression line (drawn with a dashed line) to predict the AP Maths score of a student who scored 30% for Maths. Comment on the possible reliability of this.

c. One of the students could be considered an outlier. Circle this data point.

d. Suppose the outlier is removed from the data set. How would this affect:

i. the correlation coefficient?

ii. the gradient of the least squares regression line?

9. Consider the numbers below:

$p + 2$ ; $p - 3$ ; $p + 5$ and $p + 4$

a. Determine the mean in terms of $p$

b. Now determine σ, the standard deviation in terms of $p$ if necessary. Give your answer to 2 d.p.

c. Without doing any further calculations give the standard deviation of:

i. $p + 6$ ; $p + 1$ ; $p + 9$ and $p + 8$

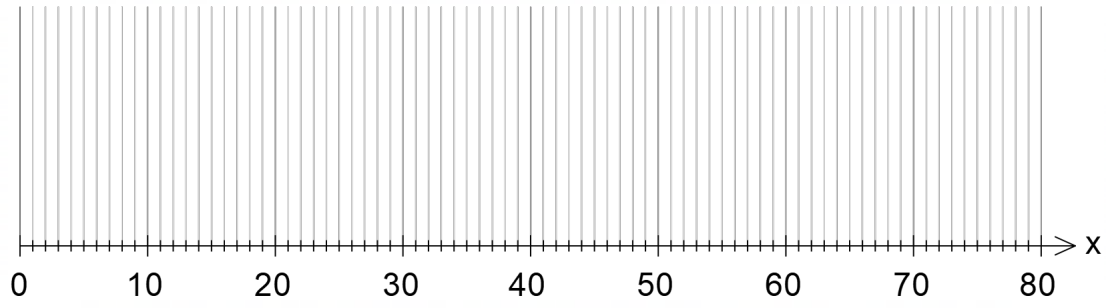
ii. $2p + 4$ ; $2p - 6$ ; $2p + 10$ and $2p + 8$

10. Consider the ranked data set below and answer the questions which follow:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | 21 | 23 | 25 | 26 | 33 | 36 | 40 | 43 | 46 | 55 | 62 | 67 | 70 | 72 |

a. Determine the standard deviation to 2 decimal places.

b. Draw a box and whisker diagram on the axes below.




d. Give the inter-quartile range.

e. Describe the skewness or symmetry of the data. If skewed, give the direction in which it is skewed.

11. Consider the following data set of 19 values in ascending order:

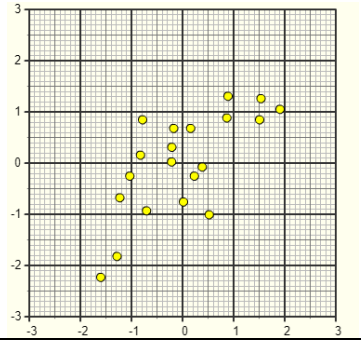
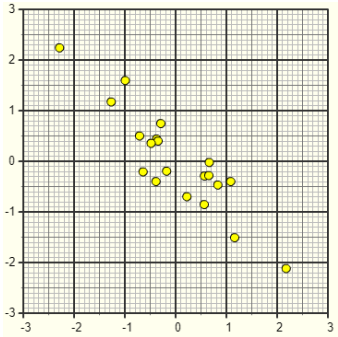
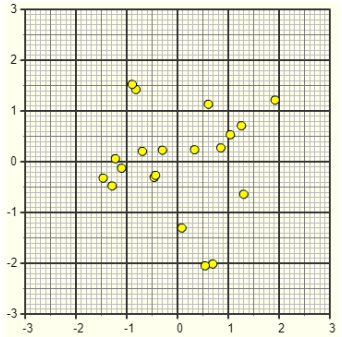
3 ; 5 ; 5 ; 6 ; 9 ; 10 ; 13 ; 17 ; 23 ; 25 ; 34 ; 41; 67 ; 68 ; 81; 83 ; 87 ; 89 ; 96

a. What is the modal value?

b. What is the inter-quartile range?

12. A data set has $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 300$ . If $\sigma = \frac{5\sqrt{2}}{2}$ then determine $n$.

13. Match the following data plots to their correlation coefficients:

| PLOT 1 | PLOT 2 | PLOT 3 |
|---|---|---|
|  |  |  |
| $r_1 = 0$ | $r_2 = 0.6$ | $r_3 = -0.9$ |

PLOT 1 _____

PLOT 2 _____

PLOT 3 _____

14. Two teachers were looking at the correlation between performance in Maths and in Geography. They used the following data.

| Maths (x) | Geography (y) |
|---|---|
| 60 | 67 |
| 69 | 80 |
| 51 | 63 |
| 62 | 72 |
| 70 | 79 |
| 63 | 69 |
| 54 | 60 |
| 65 | 74 |
| 69 | 80 |
| 57 | 68 |

a. Determine the equation of the least squares regression line giving a and b to 3 d.p. each

b. Show that $(\overline{x}\ ;\ \overline{y})$ lies on the line you found in a.

c. Comment on the nature of the relationship between the two data sets with reference to the correlation coefficient.

d. They wish to use the least squares regression line to predict a Geography result for a boy who scores 45% for Maths. Comment on the possible reliability of this

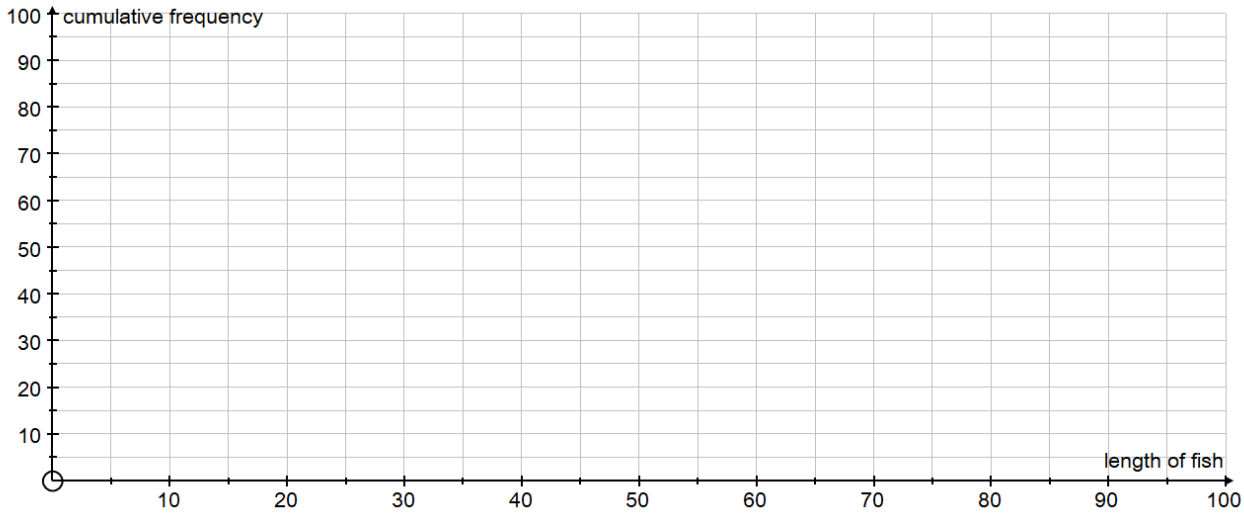
15. A summary of the lengths of fish (in cm) caught in a competition is given below.

| Length | Frequency | Cumulative frequency |
|---|---|---|
| $[0;10)$ | 4 | 4 |
| $[10;20)$ | 10 | 14 |
| $[20;30)$ | 15 | |
| $[30;40)$ | 25 | 54 |
| $[40;50)$ | 20 | 74 |
| $[50;60)$ | | 86 |
| $[60;70)$ | 8 | 94 |
| $[70;80)$ | 5 | 99 |
| $[80;90)$ | 1 | 100 |

a. Fill in the two missing entries in the above table

b. Calculate an estimate of the mean and explain why it is an estimate.

c. Draw the frequency cumulative diagram on the axes below




16. The mean of *n* numbers is m and the standard deviation is *s.*

Give the new mean and standard deviation in terms of *m* and *s* if:

a. 7 is added to each number

b. Each number is tripled

17. A teacher was looking at the correlation between the Maths (x) and Science (y) results of 40 students. Unfortunately she spilt tea on her working, leaving only the following information:

slope of regression line = 0.841
mean of maths marks = 67%
top maths mark = 89%

$$\sum_{i=1}^{40} y_i = 2800$$

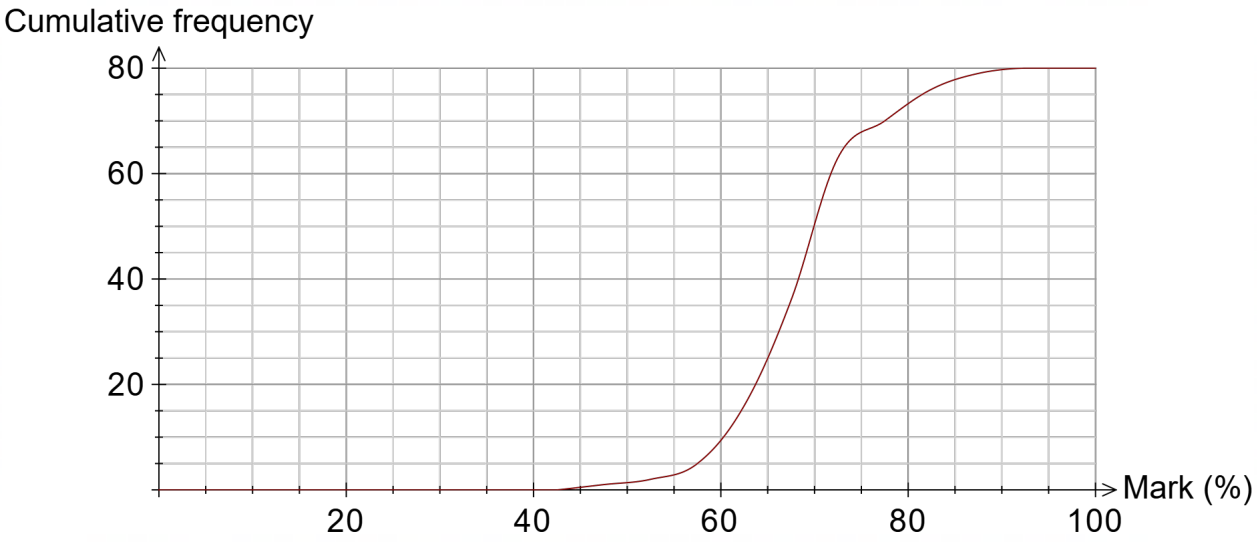
$r = 0.961$

a. Give the equation of the least squares regression line showing all working.

b. Comment on the reliability of using this line to predict the Science result of a student who scores 95% for Mathematics.

c. Comment, with justification, on correlation between Maths and Science results.

18. Consider the cumulative frequency curve below showing the distribution of marks on a test.




a. How many students wrote the test?

b. Use the curve to estimate the interquartile range showing where you have taken your readings.

c. It is decided to award a MERIT CERTIFICATE to the top 12,5% of learners. What mark is required to earn a MERIT CERTIFICATE?

19. There are 20 people in a room. The average of their ages is 32 and the standard deviation is $2\sqrt{3}$. A 28 year old and a 36 year old leave the room. What is the standard deviation of the people remaining? Give your answer to 3 d.p.