**A113 – Mathematics**

**Worksheet for Problem 09: Player Selection**

**Importance of organising data**

1. In today’s problem, a set of data is in the file, **Player\_Score.xlsx** attached along with the problem statement.
2. Discuss the information that is in the data file.

Player score, number of data values, number of days, number of players and who are the players.

1. What was the duration of the selection trials?

10 Days

1. Discuss how your team plans to use the data in it to respond to the problem.
2. Would it be possible to use the data in Player\_Score.xlsx immediately to respond to Q1(c)? Why?

No. It is raw data.

**Sorting data**

There are many ways and tools to organise data. Using a table is one of them, study Table 1 below.

1. In your Player\_Score.xlsx, try organising the data for the various players respectively similar to the table below. You may refer to the attached file on “Sorting and Filtering in MS Excel” to help you organise the raw data so that all of Andrew’s scores are listed first, followed by all of Bernard’s scores, then Calvin’s and Dan’s scores.

 Table 1: Organise by Players

|  |  |
| --- | --- |
| Player | Score |
| Andrew | 325 |
| … | … |
| Bernard | 349 |
| … | … |
| Calvin | 323 |
| … | … |
| Dan | 338 |
| … | … |

**Organising Data into a frequency table**

1. **Player\_Score.xlsx** contains the scores of Andrew, Bernard, Calvin and Dan in the selection trials. You have now arranged the scores according to the player. You can copy Andrew’s day and scores and paste them into the sheet named Andrew in Excel, do the same for the remaining players.
2. One way to display the scores for easier interpretation is to use a **frequency table**. It is a record of **how often** each value or a set of values occurs.

An example of a frequency table is shown in Table 2 below.

Table 2

|  |  |
| --- | --- |
| Score Range | Andrew’s Frequency |
| 311 to 315 | 4 |
| 316 to 320 | 16 |
| 321 to 325 | 25 |
| 326 to 330 |  |
| 331 to 335 |  |
| 336 to 340 |  |
| 341 to 345 |  |
| 346 to 350 |  |
| 351 to 355 |  |
| 356 to 360 |  |

Tabulate the scores of Andrew, Bernard, Calvin and Dan in a similar way as shown in Table 2.

1. In part (a), we have successfully organised the data into a table. What is/are the benefit(s) of organising the data using tables?
2. Relative frequency can also be interpreted as the probability of the occurrence of a particular event.



Calculate the relative frequency for Andrew’s score to fall in the range of 311 to 315.

**Organising data using charts**

Apart from using tables, data can also be organised and presented using charts. Some of the more common ones include: bar chart, pie chart, line graph, radar chart and box-and-whisker plot.

1. We can represent information from the frequency table in a graphical manner with the use of a **histogram**.
   1. Describe some benefits of representing data in a graphical manner.
   2. Using MS Excel and the frequency tables in Q4, plot the histograms of the scores of Andrew, Bernard, Calvin and Dan.
   3. We can also represent data in a graphical manner using other charts, such as pie charts. Draw a pie chart to represent just the scores of Andrew.
   4. Which of the two charts that you have drawn (histogram or pie chart) is suitable for representing Andrew’s frequency table? Explain your choice.

**Analysis of Data based on accuracy of score**

1. Let us look at the scores of Andrew, Bernard, Calvin and Dan.
   1. Determine the individual **mean, mode and median** scores for each of the four players.



* 1. Which score (mean, mode or median) is the most representative of the players’ performance? Explain.
  2. Rank the players according to the mean of the score.
  3. Determine which player (Andrew, Bernard, Calvin or Dan) is the most suitable based on the mean of the score.
  4. Would the mean, mode or median score alone be sufficient for you to justify which player you would choose to represent the school? Why?

**Analysis of Data based on variance of score**

1. We should also consider the variance or “spread” of the scores of the players when choosing the best player. Study the histograms of the scores of Andrew, Bernard, Calvin and Dan that you have plotted in Q5. Based on the shape of the histograms, whose score do you think has the least spread? Why?
2. A common tool to measure the “spread” of the data from its mean value is variance. Find out more about variance from your pre-reading materials.
   1. Using all the data in **Player\_Score.xlsx**, calculate the variance of Andrew’s scores to verify your answer in Q7. You may use the example in Table 3 to help you.

Table 3

|  |  |  |  |
| --- | --- | --- | --- |
| S/N | Score | Difference from mean, *μ* | Square of difference from mean |
| 1 | 349 | 349 - *μ* = - 0.4 | (349 - *μ*)2 = 0.16 |
| 2 | 347 | 347 - *μ* = - 2.4 | (347 - *μ*)2 = 5.76 |
| 3 | 344 | 344 - *μ* = ? | (344 - *μ*)2 = ? |
| 4 | 351 | 351 - *μ* = ? | (351 - *μ*)2 = ? |
| 5 | 356 | 356 - *μ* = ? | (356 - *μ*)2 = ? |
| Sum | 1747 | ? | ? |
| Average | *μ* = 349.4 | ? | Variance, *σ*2 = ? |

* 1. Using the same method, proceed to calculate the variance of Bernard’s, Calvin’s and Dan’s scores to verify your answer in Q7.
  2. Rank the players according to variance of the scores.
  3. Hence, determine which player (Andrew, Bernard, Calvin or Dan) is the most suitable based on the variance of their scores.
  4. A similar measurement of the “spread” of the scores is the **standard deviation**. It is also the square-root of the variance. Calculate the standard deviation of the players’ scores.



**Other Factors**

1. Discuss and list the other factors that you may consider when selecting the player to represent the school.

**Creation of various charts**

1. Based on all of the above considerations, who would you choose to represent the school? Support your decision with evidence(s) and various charts.

Attempt to create the various charts for the data you have organised in all your tables above using MS Excel **whenever possible**.

Note: You may use these websites as a guide on how to create charts in MS Excel.

<https://support.office.com/en-US/article/Create-a-chart-from-start-to-finish-a745775f-98d9-4c63-bfa8-9c00cd03ff0c>

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Practice Questions***  *(It is essential to complete these practice questions so that you can understand the concepts of this lesson better and be more confident and competent in handling related questions.)*  Organising data   1. Organise the data in the file attached, **Data 1.xlsx**, into      1. a table 2. a bar chart 3. a pie chart 4. The table below shows the height of the boys in a class.  |  |  |  |  |  | | --- | --- | --- | --- | --- | | 148 | 146 | 152 | 160 | 142 | | 130 | 146 | 145 | 149 | 165 | | 148 | 153 | 147 | 146 | 144 | | 136 | 134 | 132 | 157 | 164 | | 140 | 151 | 156 | 145 | 135 |  1. If being tall is defined as more than 160 cm, what **percentage** of the boys is considered tall? 2. Complete the frequency table and represent the data using a bar chart.  |  |  | | --- | --- | | Height | No. of boys | | 130 to <140 cm |  | | 140 cm onwards to <150 cm |  | | 150 cm onwards to <160 cm |  | | 160 cm onwards |  |  1. Represent the data in the above table using a pie chart.   Interpreting Data   1. The figure below shows the records of a survey on students’ preferred fruits in a bar chart. 2. What is the number of students that responded to this survey that preferred Apple or Pear? 3. Which fruit is the most preferred fruit by the students? 4. Which fruit is the least preferred fruit by the students? 5. What is the **relative frequency** of students that preferred Orange? 6. What is the **percentage** of students that preferred Apple or Honeydew? 7. Given that 25 female students responded in the survey that they preferred apple, what is the number of male students that responded to this survey that they preferred apple?   Frequency Table and Histogram   1. Complete the frequency table below and plot the corresponding histogram for the following list of values:   3, 23, 16, 12, 17, 19, 15, 5, 11, 20, 12, 19, 8, 22, 6   |  |  | | --- | --- | | Data Range | Frequency | | 1 – 5 |  | | 6 – 10 |  | | 11 – 15 |  | | 16 – 20 |  | | 21 – 25 |  |   Mean, Mode, Median   1. The weights of 9 students are recorded in the table below.  |  |  | | --- | --- | | Student | Weight (kg) | | A | 39.9 | | B | 31.3 | | C | 33.9 | | D | 38.7 | | E | 39.9 | | F | 32.8 | | G | 37.9 | | H | 33.9 | | I | 36.6 |  1. What is the mean weight of these 9 students? 2. What is/are the mode(s) of the weights of these 9 students? 3. If the table is amended to include the weight of Student J, the mean weight of these 10 students will be 36.4 kg. What is the weight of Student J?   [You may want to watch the following [video](https://docs.google.com/file/d/0BxYqWZfRUFnoZW40bUtjcHhfLW8/edit?usp=sharing) which would help to recap some of the key concepts learnt before attempting the following question.]   1. The table below shows the distribution of weight of 22 students in kilograms (kg):  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 56 | 42 | 43 | 70 | 66 | 50 | 48 | 90 | 45 | 61 | 80 | | 48 | 50 | 61 | 47 | 52 | 68 | 74 | 58 | 49 | 45 | 82 |  * 1. What is the average for all the students in this group?   2. State the modal weight for this group of students. (There can be more than one mode)      * 1. What is the median weight for this group of students?   2. Given that one additional student was added to this group of students and the average weight becomes 58.5kg, determine the weight of this additional student.  1. The table below shows the distribution of the number of hours spent surfing the internet by a group of teenagers in six months.  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Number of hours surfing the internet | 450 | 650 | 700 | 740 | 800 | 850 | 900 | 1000 | | Frequency | 5 | 11 | 14 | 18 | 19 | 13 | 10 | 4 |  1. State the median number of hours spent surfing the internet. 2. What is the modal number of hours spent surfing the internet? 3. What is the total number of hours that teenagers spent surfing the internet for **less than 740 hours**? 4. Calculate the mean number of hours spent by teenagers when they surf the internet in six months. /// 5. The histogram below shows the income for part-time work done by 100 students.   Make use of the bar chart to answer the following questions:   1. What is the relative frequency of students who earned $230? 2. State the mode(s) of the income for this group of students. 3. What is the median income for this group of students? 4. Calculate the mean income for this group of students.   Standard Deviation   1. Determine the mean and standard deviation for the following set of data:   3, 23, 16, 12, 17  Mean =  Standard deviation = |

**Exploring further**

1. Calculate the range of each player’s scores and tabulate them in Table 4, together with the standard deviation of each player’s score.

(Hint: The range is the difference between each player’s highest and lowest score.)

Table 4

|  |  |  |
| --- | --- | --- |
| Player | Range | Standard Deviation |
| Andrew |  |  |
| Bernard |  |  |
| Calvin |  |  |
| Dan |  |  |

Based on the comparison above, is the range, which is easier to calculate, a good replacement for the standard deviation?

1. A fifth player, Edward, was included in the trials. His scores have a range of 35. Details of Edward’s scores can be found in this document.



The coach commented that Edward’s performance was comparable to that of Calvin. Is the coach’s comment justified?