NASCA Quantitative Literacy Materials Draft 1

# Topic 4: Data Handling

# Topic 4: Data Handling

#### **Learning Outcomes**

By the end of the unit you should be able to:

* Identify samples and populations.
* Use appropriate statistical methods for collecting, classifying, summarizing, representing and interpreting data.
* Compare data from different sources and samples.

### **Introduction**

The principles underlying this topic is primarily to get you to engage with the data and critically analyse and then communicate with data.

Although some experience in collecting, organising and interpreting data maybe useful, the focus in this topic is more on interpreting data rather than gathering or generating it.

To develop an analytical and critical approach towards arguments based on data, you ought to know that data can be represented/misrepresented and interpreted/misinterpreted in different ways.

## Samples and Populations

* A Sample is when we collect data just for selected members of the group.
* A Population is when we collect data for every member of the group, eg. Census, the whole population is involved.

**Illustrative example** showing differences:

It is not practical to count all the damaged oranges in an orchard (population), but it is possible to count the number damaged in a set of oranges (sample) taken from the orchard (population).

**Illustrative example:**

Before an election, a news agency wanted to conduct a poll to determine who would win.

Which of the following represents a sample and which a population?

1. A selection of voters of different ages
2. All voters

**Solution:** *(a) Sample (b) Population*

**Illustrative example:**

A popular artist wanted to know what people thought about his painting. Which of the following will represent a sample and which a population?

1. Every person who purchased the painting
2. A selection of people who purchased the painting

**Solution:** *(a) Population (b) Sample*

**Activity for reflection.**

**What you need:**

* **You need to recall definitions of sample and population.**

**Practice exercise 1**

1. In each of the following statements, identify the sample and the population.
	1. A city councilor wanted to know how her constituents felt about a planned stadium in their city. She randomly selected 100 names from the city phone directory and conducted a phone survey.

**Guided reflection:**

**Does the selection involve all the names of people in the directory?**

**Does the selection involve a small collection of people?**

* 1. A school teacher randomly selected 10 parents of his learners. He then surveyed the parents about their opinions of different resource materials.

**Guided reflection:**

**Does the selected parents represent the entire school?**

* 1. The municipality in a certain city wants to know about the thickness of tar on a busy road. They randomly select 30 points on the road and measures the thickness of tar at each of these points.

**Guided reflection:**

**Does the selected points chosen represent all possible points on the**

**road?**

* 1. SANRAL wants to know how many cars pass through a certain toll gate in a day. A camera is installed over one lane of the road which photographs the license plate of every tenth vehicle that passes through that lane.

**Guided reflection:**

**Is the camera able to capture all the cars in every lane?**

* 1. A farmer wants to know how many oranges are damaged in his orchard. He collects 50 boxes of oranges and inspects all of them.

**Guided reflection:**

**Is the farmer able to identify every damaged orange in his entire**

**orchard?**

## Quantitative and Qualitative data

There are two types of data:

**Quantitative data**

* Is numerical information (numbers)
* Is made up two types: Discrete (countable items – whole numbers) and Continuous (measurements – decimals, fractions)

 Examples of discrete data: Number of products in a brochure; Number of employees you have.

 Examples of continuous data: Dimensions of a specific product. Time taken to reach your

 destination.

**Qualitative data**

* Descriptive information, describing something.
* Are measures of categories or types and may be represented by name, symbol, category or type.
* Also referred to as categorical data. Eg. Favourite colour or pet.

**Discrete** data is counted while **continuous** data is measured.

**Quantitative** data is numerical information while **Qualitative** data is descriptive information.

**Illustrative Examples**

Classify the following as quantitative or qualitative data. If the data is quantitative, state whether it is discrete or continuous data.

1. How many pets do you have?
2. In which city is the school located?
3. How do you travel to work?
4. How much do you earn?
5. Which brand of cool drink is the most popular?
6. Which country won the last soccer world cup trophy?

**Solutions:**

1. quantitative data (discrete) (b) qualitative data (c) qualitative data

(d) quantitative data (continuous) (e) qualitative data (f) qualitative data

Numerical Data and Categorical Data

**Numerical data** – data that have meaning as a measurement or a count, for example a person’s height, weight, blood pressure or number of pages in a book or number of teeth a human being has. Numerical data can be either discrete (counted as a whole number) or continuous (measurements that continue, decimals). Numerical data is quantitative.

**Categorical data** – data that represent characteristics such as a person’s gender, marital status, favourite movies, hometown, etc. Categorical data are qualitative data.

**Activity for guided reflection:**

**Practice Exercise 2**

1. Choose whether the following statements represent qualitative or quantitative data.

**Guided reflection:**

**Remember if you are able to quantify (state in form of numbers) then this will represent quantitative data and if you are able to describe the data without numbers it is qualitative data.**

1.1 My friend is very depressed.

1.2 The newly born baby weight 3,5 kg at birth.

1.3 The sea is greyish blue.

1.4 There are 32 countries participating in the 2018 soccer world cup.

1.5 Ashley has R280.

1. Determine whether the following statements are categorical or numerical.

**Guided reflection:**

**Remember numerical data is quantitative data that can be discrete, counted as a whole number or continuous data such as decimals.**

**Categorical data is qualitative data descriptive characteristics.**

* 1. The following test scores were achieved by the top learners in a class test: 99; 88; 86;

72; 70; 70 and 65.

* 1. A car dealership did a stock count to find out how many of the following cars were on

 the floor: Toyota; Mercedes; BMW; VW and Hyundai.

* 1. A snap survey was taken at school to determine the type of backpacks carried to school:

Five students had black backpacks; ten had blue backpacks and four had green backpacks.

1. Determine whether the following data sets are discrete or continuous.

**Guided reflection:**

**Remember that discrete data can be counted using whole numbers, like how many chairs, tables, people, etc. whereas continuous data is measurements given as part of a whole, example 35,5 kg.**

* 1. The number of shoes in your shoe rack.
	2. The weights of people in a lift.
	3. The age of a person.
	4. The height of players in a netball team.
	5. The number of words in a book.

## Data collection methods – selected types.

Data can be collected using some of following methods:

* Surveys and Questionnaires
* Interviews and focus groups
* Direct observations
* Document analysis, registers or registrations

Surveys and Questionnaires

**Purpose:**

Surveys and questionnaires are useful when gathering specific information from participants.

**Advantages:**

* One is able to gather information from large groups of people (high representativeness)
* Cost effective – low costs involved and sometimes even can be free (goggles “survey monkey”)
* Convenient method – participants can respond electronically (online) at their convenience and able to target global participants.

**Disadvantages:**

* Information can lack detail.
* Inflexible – the type of questions are sometimes rigid and does not allow for changes.
* Not ideal for controversial issues – participants may be apprehensive to answer honestly – afraid of victimization.

Interviews and Focus groups

**Purpose:**

Useful in collecting detailed qualitative data and lived (real) experiences. Interviews can be conducted one on one or in small groups (focus groups).

**Advantages:**

* Useful in understanding peoples lived experiences in detail.
* More accurate – face to face helps get honest answers like gender, age, race, etc.
* Capture verbal and non-verbal cues – body language – indicate discomfort or enthusiasm.
* Capture emotions and behaviours.

**Disadvantages**:

* Time consuming as more time is required to conduct, transcribe and analyse.
* Costly – require staff, transport and sometimes incentives for respondents.
* Sample size – limited and depends on staff, availability of respondents and quality of interviewers.

Observations

**Purpose:**

Observations are generally unnoticeable and can be used for gathering vital information about children’s growth and development, the home situation and environment, etc.

**Advantages:**

* Simplest method – does not require equipment, technology.
* Greater accuracy
* Useful for formulating hypothesis.

**Disadvantages:**

* Some occurrences may be inaccessible – not open for observation.
* Some occurrences which are accessible may not actually happen – uncertainty of event.
* Lack of reliability.
* Personal bias
* Expensive.

Document analysis, registers or registrations

**Purpose:**

One can get access to information that would be difficult to get in any other way.

**Advantages:**

* Access to inaccessible subjects or data
* Large sample size possible.
* Relatively low cost
* Many documents are good quality and contains valuable information relating to past and present.

**Disadvantages:**

* Can be viewed as too subjective
* Time consuming
* Depends on the role of the researcher
* Documents are not usually designed with research in mind – information may be incomplete.

**Activity for guided reflection**

**Practice Exercise 3**

1. When collecting data, why is it advisable to firstly conduct a sample survey rather than using any other method of data collection?

**Guided reflection:**

**Imagine conducting an interview with someone who knows nothing about your research study or has nothing substantial to contribute to the data you require.**

1. List some of the criteria you would consider when choosing a data collection method.

**Guided reflection:**

**Remember you want to collect data efficiently and effectively. What are some of the characteristics that will ensure these two qualities are met?**

1. Consider the following data collection methods:
2. Surveys/questionnaires
3. Observation
4. Interview/focus group
5. Document analysis

Choose which data collection method(s) listed above you would most likely employ to gather data for the following topics:

* 1. A study of the salaries of teachers in a particular province.
	2. To determine if a politician is popular in his constituency.
	3. The ages of people frequenting a mall.
	4. The average height of students in your class
	5. The favourite dish chosen by customers at a restaurant.
	6. The time your mum spends doing household chores each week.
	7. Which sporting hero made the most money playing sport.

**Solutions for Practice exercises.**

**Practice exercise 1**

* 1. The population is everyone listed in the city phone directory; the sample is the 100 people selected.
	2. The sample is the selected parents while the population is all the parents.
	3. The population is every point on the road while the sample is the 30 selected points.
	4. The population is all the vehicles that pass through the lane while the sample is the group of every tenth vehicle that pass through the lane.
	5. The sample is the 50 boxes collected while the population is the orchard of oranges.

**Practice exercise 2**

1.1 Qualitative

1.2 Quantitative

1.3 Qualitative

1.4 Quantitative

1.5 Quantitative

2.1 Numerical

2.2 Categorical

2,3 Categorical

3.1 Discrete

3.2 Continuous

3.3 Continuous

3.4 Continuous

3.5 Discrete

**Practice exercise 3**

1. It is less expensive; quicker to conduct; serves specialized needs; more convenient; eliminates unqualified respondents – hones in on target group.
2. Cost (budget); Time; Sample size; personnel required.
3. 3.1 Document analysis or survey (survey could elicit dishonest responses especially salaries)
	1. Interview
	2. Survey or questionnaire
	3. Survey
	4. Observation or survey
	5. Observation
	6. Document analysis

## UNIT 2 – Measures of Central Tendency

#### **Learning Outcomes**

By the end of the unit you should be able to:

* Understand and explain the three measures of central tendencies.
* Complete calculations involving the three measures of central tendencies.

#### **Introduction**

A measure of central tendency is a single value that tries to define a set of data by finding the middle position within that set of data.

* The mean, median and mode are valid measures of central tendencies.
* The **mean** is the average and is calculated by finding the sum of the data values and then dividing by the number of data values.
* The **median** is the middle value or the mean of two middle values when the data is arranged in numerical order.
* The **mode** is the value that appears the most number of times. It is possible to have more than one mode.
* An **outlier** is a value that "lies outside”, that is, it is much smaller or larger than most of the other values in a set of data.
* Use **mean** to describe a set of data that **does not** have an outlier.
* Use **median** to describe a set of data that **has** an outlier.
* Use **mode** when asked to choose most **popular** item in a set of data.

**Purpose:**

The purpose of measures of central tendency is for you to identify the position of the centre of data sets.

**Mode:** This is used when: – the observation that is most frequently observed is desired; when a quick estimate of central tendency is desired or when the data is categorical.

**Median:** This is appropriate when: – the centre/middle value of the data set is desired; when you need to determine whether additional data points fall either above or below the midpoint; the data is highly skewed; outliers exist which will affect the mean.

**Mean:** This is used when:– the data is symmetrical or not really skewed.

**Illustrative Examples**

Pearl scored the following percentages for her first 5 History tests: 100; 72; 86; 72 and 92.

Find the mean, median and mode for her test scores.

Mean = $ \frac{Sum of scores}{ Number of scores}$

 =$ \frac{100+72+86+72+92}{ 5}$

 = 84,4%

Median: First arrange the marks in order of magnitude (either from smallest to biggest or biggest to

 smallest).

 Arranged data: 72; 72; 86; 92; 100.

 Therefore the median will be the middle value which is 86.

Mode: Data value that occurs most number of times: Therefore the mode is: 72.

**Activity for guided reflection**

**Practice exercise 4**

1. Calculate the mean, median and mode for the following data set:

5; 15; 10; 15; 5; 10; 10; 20; 25; 15; 20; 15.

**Guided reflection:**

**Recall the definitions of the concepts and the method involved in calculating these central measures of tendencies:**

**Mean: Sum total divided by number of scores; Median is the middle value in arranged data, not forgetting that an even set of data has two middle values which must be added and divided by 2 and mode is the data value repeating itself the most number of times.**

1. For his first 6 English tests, Alfie received the following percentage scores:

72; 68; 95; 92; 77; ?. Alfie somehow misplaced his last test (?) but he did remember that his mean score for all 6 tests was 80.

Help Alfie figure out what his actual percentage score for his last test was.

**Guided reflection:**

**This is an example of reverse calculation, that is, instead of adding the scores and dividing by the total number we do the reverse, that is, we multiply and subtract the scores we have.**

1. The following scores are arranged in ascending order. If the median of the data is 25, find the value of x.

 17; x; 24; x + 7; 35; 36; 46.

**Guided reflection:**

**Remember median is the middle score. How many scores do we have? Also note we know the middle score.**

**Solution – Practice exercise 4**

1. Mean = $\frac{5+15+10+15+5+10+10+20+25+15+20+15}{ 12}$

 =$ \frac{165}{ 12}$

 = 13,75

Median: 5; 5; 10; 10; 10; 15; 15; 15; 15; 20; 20; 25

Therefore the median is the average of the two middle values: (15 + 15) $÷$ 2 = 15.

 Mode: 15

1. We take the mean and multiply by the number of tests: 80 $×$ 6 = 480.

We then take 480 and subtract the sum total of the five tests: 480 – (72+68+95+92+77) = 76.

Therefore, Alfie’s 6th test score was 76. One can always double check your answer by reworking the mean with 76 as the 6th test score and see whether you get to a mean of 80.

1. The median is 25, therefore x + 7 = 25, implying x = 18.

## UNIT 3 – Measures of Spread

#### **Learning Outcomes**

By the end of the unit you should be able to:

* Understand and explain the concept of measures of spread.
* Understand how to recognize quartiles and interpret diagrams involving quartiles.
* Complete calculations involving quartiles.
* Interpret box-and-whisker plots/diagrams.

**Introduction**

* A measure of spread refers to how the data within the set is spread out, or dispersed or scattered about the mean.
* Sometimes a data set may have the same mean and median. Therefore to determine how these data sets are different requires us to calculate the spread of the data set, that is, how is the data spread out?
* We can use the range and the Interquartile range to measure the spread.
* Range is the difference between the largest data value and the smallest data value in the data set.
* The interquartile range is another form of range which divides the data set into four equal parts (quarters). The three values that form the four divisions are called quartiles: First quartile (Lower quartile or Q1), Second quartile (Median or Q2) and Third quartile (Upper quartile or Q3).
* The Interquartile Range (IQR) is the difference between the third quartile and the first quartile

 (IQR = Q3 – Q1).

* The IQR represents 50% of the data which eliminates the influence of any outliers.
* Percentiles divide a data set into hundred parts. Commonly used percentiles are the 25th (quartile 1), 50th (median or quartile 2) and 75th (quartile 3).

**Purpose:**

Summarising a dataset can help you understand the data, especially when the dataset is large. As discussed in the [Measures of Central Tendency](http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical%2Blanguage%2B-%2Bstatistical%2Blanguage%2Bglossary#Measures%20of%20central%20tendency) section , the mode, median, and mean summarises the data into a single value that is representative of all the values in the dataset, but this is only one part of the 'picture' that summarises a dataset. Measures of spread summarise the data in a way that shows you how scattered (spread out) the values are and how much they differ from the mean value.

**Illustrative Examples**

1. Determine the range, quartiles (lower, middle and upper) and interquartile range for the

 following set of data: 5; 9; 8; 4; 3; 8; 6

 **Solution:** Arrange the data set in order; then divide them into quarters.

 3 ④ 5 ⑥ 8 ⑧ 9

 **Q1 Q2 Q3**

Range = Highest value – Lowest value = 9 – 3 = 6

Quartile 1 (Lower quartile) = 4

 Quartile 2 (Median) = 6

 Quartile 3 (Upper quartile) = 8

 Interquartile range = Q3 – Q1 = 8 – 4 = 4

1. Determine the quartiles, for the following set of data: 5; 9; 8; 4; 3; 9; 6; 6.

**Solution:**

**Step 1: Arrange the data in ascending order.**

**Step 2: Locate the median – the middle value: Since there is an even number of values we locate**

 **the median (Q2) between the two middle values, namely (6 + 6)** $÷$ **2 = 6**

**Step 3: Locate the median of the first half of the data values, this is Q1 = (4 + 5)** $÷$ **2 = 4,5**

**Step 4: Locate the median of the second half of the data values, this is Q3 = (8 + 9)** $÷$ **2 = 8,5**

**See illustrations below for calculations.**

 3 4 5 6 6 8 9 9

 **Q2 = (6+6) ÷ 2 = 6**

 **Q1 = (4+5) ÷ 2 = 4,5 Q3 = (8+9) ÷ 2 = 8,5**

## UNIT 4 – Representation of data

#### **Learning Outcomes**

By the end of the unit you should be able to:

* Identify different forms of data representation.
* Represent data using the different forms of representation, namely frequency tables, box and whisker plots, scatter plots, bar graphs, histograms, line graphs, pie chart, stem and leaf diagram and pictograms.

**Introduction**

* Ungrouped data can be represented by means of a **frequency table.**
* Data can be represented using box-and whisker plots, scatter plots, bar graphs, pie charts, histograms, straight line graphs, broken line graphs, pictographs,
* Tally marks are represented by a tally mark (/) each time a data value appears. When a data value is repeated 5 times, then 4 tally marks are used with the fifth drawn across, eg. 卌.

**Illustrative Examples**

1. A 6-sided dice was thrown 30 times and the following outcomes were observed:

 1, 4, 2, 4, 6, 1, 2, 3, 6, 5, 4, 4, 3, 1, 1, 3, 1, 1, 5, 6, 6, 2, 2, 3, 4, 2, 5, 5, 6, 2

 Draw a frequency table to represent the above data.

 **Solution:** **Frequency table**

|  |  |  |
| --- | --- | --- |
| **NUMBER** | **TALLY** | **FREQUENCY** |
| **1** | **//// /** | 6 |
| **2** | **//// /** | 6 |
| **3** | **////** | 4 |
| **4** | **////** | 5 |
| **5** | **////** | 4 |
| **6** | **////** | 5 |

1. The scores obtained by 40 students in an examination are given below:

8, 47, 22, 31, 17, 13, 38, 26,

3, 34, 29, 11, 22, 7, 15, 24,

38, 31, 21, 35, 42, 24, 45, 23,

21, 27, 29, 49, 25, 48, 21, 15,

18, 27, 19, 45, 14, 34, 37, 34

Draw up a frequency table to group the data using 5 class intervals, 0 -10, where 10 is excluded.

**Solution:**

|  |  |
| --- | --- |
| **Scores** | **Frequency** |
| **0 – 10** | **3** |
| **10 – 20** | **8** |
| **20 – 30** | **14** |
| **30 - 40** | **9** |
| **40 – 50** | **6** |

The Box-and-whisker plot

* You will not be required to draw the box-and-whisker plot, but rather you will be assessed on the interpretation of it.
* The box-and-whisker plot is a simple way of representing statistical data on a number line in which a rectangle is drawn to represent the second and third quartiles, usually with a vertical line inside to indicate the median value. The lower and upper quartiles are shown as horizontal lines on either side of the rectangle. See example below:



The above box-and-whisker shows a five number summary as follows:

1. The minimum value of 3
2. Q1 = Quartile 1 which is 7
3. Q2 = Quartile 2 which is the median 13
4. Q3 = Quartile 3 which is 15
5. Maximum value is 21
* The 5 number summary of a set of observable data consists of the following:

Maximum– the largest observation

Upper Quartile (Q3) – a value that separates the largest 25% of the observations from the smallest 75%

Median (Q2) – a value that separates the largest 50% of the observations from the smallest 50%

Lower Quartile (Q1) – a value that separates the largest 75% of the observations from the smallest 25%.

Minimum – the smallest observation.

**Purpose:**

A Box and Whisker Plot is a convenient means you can use to visually present the data distribution through their quartiles. The lines extending parallel from the boxes are known as the “whiskers”, which are used to indicate variability outside the upper and lower quartiles.

**Practice exercise 5**

1. Refer to the following box-and-whisker plots which represents two tests written by a grade 12 class and then answer the questions which follow:



* 1. Determine the 5 number summary for each of the tests.

**Guided reflection:**

**Remember the five number summaries shows us the spread of the data in quarters.**

* 1. State with reasons in which test the students performed better.

**Guided reflection:**

**Calculate the IQR to check the spread of marks: Check which test has more scores on the right of the median.**

* 1. Determine the Interquartile range for each test.

**Guided reflection:**

**Recall the formula for IQR = Upper quartile value minus lower quartile value.**

**Solutions to Practice Exercise 5**

 1.1

|  |  |  |
| --- | --- | --- |
| **5 NUMBER SUMMARY** | **TEST A** | **TEST B** |
| MINIMUM VALUE | 3 | 3 |
| LOWER QUARTILE (Q1) | 9 | 6 |
| MEDIAN (Q2) | 12 | 12 |
| UPPER QUARTILE (Q3) | 16 | 16 |
| MAXIMUM VALUE | 21 | 21 |

 1.2 TEST A – More students had scores to the right of the median. The scores obtained in test A are

 consistent as they are clustered close together around the median.

* 1. IQR Test A = Q3 – Q1 = 16 – 9 = 7 andIQR Test B = Q3 – Q1 = 16 – 6 = 10

**Activity for guided reflection**

**Practice exercise 6**

1. The following data shows the test scores out of 10 of a group of 30 students.

 2; 3; 3; 4; 8; 7; 9; 2; 3; 1

 5; 6; 4; 8; 5; 1; 2; 9; 5; 5

 5; 4; 6; 10; 9; 6; 7; 6; 5; 6.

* 1. Draw a frequency table to represent the above data.

**Guided reflection:**

**Remember a frequency table is a way of writing information in a way that is easy to read. You need to show three columns: The scores; Tally marks: Frequency column.**

* 1. Determine the modal test mark.

**Guided reflection:**

**Remember the definition for mode.**

* 1. Determine the mean test mark.

**Guided reflection:**

**Recall the formula for calculation of mean.**

* 1. Write down the median test mark.

**Guided reflection:**

**Recall the position of the median score.**

* 1. Calculate the range for the data set.

**Guided reflection:**

**Recall the formula for range.**

* 1. Determine the 5 number summaries for the above data.

**Guided reflection:**

**Remember the 5 number summaries can only be determined once the data is arranged in ascending order.**

* 1. Comment on the skewness of the data.

**Guided reflection:**

**Recall that the box-and-whisker plot**

**Solution – Practice Exercise 6**

 1.1

|  |  |  |
| --- | --- | --- |
| **Scores** | **Tally** | **Frequency** |
| 1 | // | 2 |
| 2 | /// | 3 |
| 3 | /// | 3 |
| 4 | /// | 3 |
| 5 | //// / | 6 |
| 6 | //// | 5 |
| 7 | // | 2 |
| 8 | // | 2 |
| 9 | /// | 3 |
| 10 | / | 1 |

 1.2 Score occurring most frequently: 5

 1.3 Mean = $\frac{\left(1×2\right)+\left(2×3\right)+\left(3×3\right)+\left(4×3\right)+\left(5×6\right)+\left(6×5\right)+\left(7×2\right)+\left(8×2\right)+\left(9×3\right)+10}{ 30}$

 =$ \frac{156}{ 30}$

 = 5,2

* 1. Median will be located between the 15th and 16th score

 Therefore, median will be the average of (5 + 5) $÷$ 2 = 5.

* 1. Range = Highest score – Lowest score

 = 10 – 1 = 9

* 1. Arrange the data in descending order:

 1; 1; 2; 2; 2; 3; 3; 3; 4; 4; 4; 5; 5; 5; 5; 5; 5; 6; 6; 6; 6; 6; 7; 7;

 8; 8; 9; 9; 9; 10

Divide the data into quarters: Shading shows us they are exactly 7 values in each quarter.

Therefore the 5 number summaries are:

Minimum value = 1

Lower quartile = 3

Median = (5+5) $÷$2 = 5

Upper quartile = 7

Maximum value = 10

* 1. IQR = Upper quartile – Lower quartile = 7 – 3 = 4

Q3 – Q2 = 7 – 5 = 2

Q2 – Q1 = 5 – 3 = 2

Therefore the data represented shows no skewness as the box-and-whisker plot would be symmetrical, meaning the box on either side of the median will be the same width.

Scatter Plots

* You will **NOT** be required to draw scatter plots but you will be assessed on the interpretation.
* A scatter plot is used to display two sets of data in order to find a relationship between them.
* A scatter plot shows trends.
* A scatter plot shows a positive trend if, as one set of data values increases, the other set tends to increase.
* A scatter plot shows a negative trend if, as one data set of values increases, the other set tends to decrease.
* A scatter plot shows no trend if the ordered pairs show no correlation.
* Correlation is made of co- ("together"), and relation – thus relationship.

**Purpose:**

Scatter plots show you how much one variable is affected by another. It shows you the relationship between two variables which is called their correlation.

The four diagrams that follow show four different possible scatter plots with a “line of best fit”. A line of best fit (or "trend" line) is a straight line that best represents the data on a scatter plot. This line may pass through some of the points, none of the points, or all of the points.

If this line of best fit has a positive gradient (going uphill) then the correlation is positive and if the line of best fit has a negative gradient (going downhill) then the correlation is negative. If the points lie close to each other then they will have a strong correlation and likewise if they lie far apart from each other it will be weak or sometimes no correlation.

Sometimes there might be one or two points that lie far away from most of the other points. These points are called outliers.

**Scatter plots showing correlations.**



**Illustrative Example**

The following data represents the amount of hours spent studying for a test and the corresponding test scores as a percentage.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Hours studied** | **3** | **5** | **2** | **6** | **7** | **1** | **2** | **7** | **1** | **7** |
| **Test scores (%)** | **80** | **90** | **75** | **80** | **90** | **50** | **65** | **85** | **40** | **100** |

 The scatter plot representing the above data is drawn below:

The above scatter plot shows a strong positive correlation, that is the test scores depends on the number of hours studied. If you study for longer hours you performed better. You will notice that there is more or less the same number of points on either side of the line of best fit.

**Activity for guided reflection**

**Practice Exercise 7**

Study the following scatter plot showing the relationship between age of student and the amount of time spent studying and then answer the questions that follow:

1. Does the above scatter plot show a positive, negative or no correlation?

**Guided reflection:**

**Remember the line of best fit is an indicator of the correlation.**

1. Write down the coordinates of the outlier.

**Guided reflection:**

**Remember the scatter plot represents 2 variables, age (horizontal axis) and time (vertical axis)**

**Coordinates mean the 2 variables in order, first the horizontal axis value and then the vertical axis value.**

1. Comment on the relationship between age and the amount of time spent studying.

**Solution for Practice Exercise 7**

1. Positive correlation. Straight line is going uphill.
2. Coordinates (age and time) is (12: 71)
3. The older you get the more time you spend studying.

**More different types of data representations**

In this unit you will be introduced to more different types of ways to represent data.

**Purpose:**

We are going to look at more different ways to represent data and to compare these so that we get the best form of representation. We will look at bar graphs, histograms, line graphs, pie charts and pictographs.

**Bar graphs (horizontal and vertical)**

 Useful to compare categories (qualitative data where the independent variable is non-numerical)

 and grouped discrete quantitative data (scores in a test). Bar graphs have gaps between bars.

 **Illustrative example: Horizontal bar graph Vertical bar graph**

|  |  |
| --- | --- |
|  |  |

**Histogram**

Useful to represent grouped continuous data. Histograms have no gaps between bars, it continues.

**Illustrative example:**

Following data shows the number of different learners going on holiday from 1970 to 2020

|  |  |
| --- | --- |
| 1970 – 1980 | 60 |
| 1980 – 1990 | 80 |
| 1990 – 2000 | 48 |
| 2000 – 2010 | 100 |
| 2010 – 2020 | 40 |

**HISTOGRAM representing the data for students going on holiday**



**Line graphs and Broken line graphs**

A broken line graph is formed when data points are joined by line segments. The horizontal axis represents discrete quantities such as months or years, whereas the vertical axis can represent continuous quantities such as rainfall or temperature.

A continuous line graph consists of an unbroken line in which both axes represent continuous quantities such as distance and time.

**Purpose:**

Line graphs are useful as they show trends that can be easily widened. This means that with some graphs it might be possible to continue the line to show what might happen in the future.

**Illustrative examples:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| The following table represents the production of oranges in thousands of tonnes for the different years. Draw a broken line graph to represent the data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 1980 | 1990 | 2000 | 2010 | 2020 |
| Oranges (tonnes) | 36 | 28 | 69 | 74 | 58 |

 | The following table shows Juhi’s weight for the first six months of 2019. Draw a line graph to illustrate the data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Months** | Jan | Feb | Mar | Apr | May | Jun |
| **Weight (kg)** | 49 | 54 | 61 | 69 | 73 | 75 |

 |

**Pie Chart/Graph**

A pie chart is a circular diagram used to display data.

**Purpose:**

Pie charts are normally used to show different parts (sectors) and what proportion of the whole each part represents. This proportion can be written as a fraction or percentage of the whole. Ideally the number of sectors (categories) should be limited (2 to 8).

**Illustrative example**

120 students were asked what mode of transport they used to get to school. This data was used to draw a pie chart which follows the table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| MODE OF TRANSPORT | WALK | CAR | BICYCLE | BUS | TAXI |
| NUMBER OF STUDENTS | 30 | 60 | 5 | 8 | 17 |

**How to draw this pie chart:**

Remember the measure of the angle at the centre is proportional to the frequency.

The following formula can be used to obtain the angle at the centre:

$$Measure of the angle at the centre=\frac{Frequency of the category}{Total frequency}×360^{°}$$

Therefore the following calculations can be made from the table using the above formula:

Walk = $\frac{30}{120}×360^{°}=90^{°}$ Bicycle =$ \frac{5}{120}×360^{°}=15^{°}$ Taxi = $\frac{17}{120}×360^{°}=51^{°}$

Car =$\frac{60}{120}×360^{°}=180^{°}$ Bus = $ \frac{8}{120}×360^{°}=24^{°}$

**Pictograph or Pictograms**

A Pictograph is a way of showing data using pictures (images). Each picture stands for a certain number of an item.

**Purpose:**

Used to illustrate broad differences between categories.

**How to draw:**

Draw simple pictures (images or symbols) to represent the frequency. A key is to be provided to indicate the value of each picture represented.

**Illustrative example:**

A busy restaurant recorded the number of eggs ordered and consumed by its customers for a week.

Study the pictograph to determine how many trays of eggs were used for the entire week if each tray contained 36 eggs.

 PICTOGRAPH SHOWING THE NUMBER OF EGGS CONSUMED AT A BUSY RESTAURANT.

|  |
| --- |
|  |

**Solution:**

Each picture represents 20 eggs, therefore half a picture represents 10 eggs.

Therefore total number of eggs consumed = 17,5 $×$ 20 = 350 eggs.

Therefore, the number of trays used = 350 $÷$ 36 = 9,7 $≈$ 10 trays.

**Practice Exercise 8**

1. Use the following data to draw a vertical bar graph representing the number of students having a birthday in the various months of the year.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Month | J | F | M | A | M | J | J | A | S | O | N | D |
| Frequency | 33 | 3 | 21 | 18 | 6 | 30 | 24 | 18 | 54 | 21 | 9 | 6 |

1. The following histograms represent the results of a survey held in a school involving 40 boys and 46 girls. The survey was to establish how many boys and girls used social media and the amount of time in a day they spend using social media.

|  |  |
| --- | --- |
|  |  |

Adapted: <https://www.shmoop.com/basic-statistics-probability/histograms-exercises.html>. 25/8/18

Study the histograms above and answer the questions that follow.

* 1. State how many boys spent more than 90 minutes accessing social media sites?
	2. Write down how many girls spent between 75 minutes to 90 minutes accessing social media sites.
	3. Which time slot is most popular for girls to access social media sites?
	4. Give the time slot which boys find most popular to access social media.
	5. What conclusion can you make when you compare the time spent by girls to the time spent by boys accessing social media sites?
1. The following data shows the number of different vehicles that frequent a Mall during one busy weekend:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type of vehicle | Cars | Bakkies | Motorcycles | Taxis | Other |
| Number | 26 | 30 | 5 | 29 | 10 |

Study the table above and then answer the questions that follow:

* 1. Represent the above data using a pie chart.
	2. Now use a broken line graph to represent the same data.

**Solutions – Practice Exercise 8**

1.1

* 1. 8
	2. None/Nil/Zero
	3. 210 to 240 minutes
	4. 30 to 45 minutes
	5. Girls spend more time on social media sites.

3.1

3.2

**UNIT 5 – Misrepresentation of Data**

#### **Learning Outcomes**

By the end of the unit you should be able to:

* Identify different forms of misrepresentation of data.
* Give suitable reasons how and why data is sometimes misrepresented.

**Introduction**

Very often data is distorted (change the appearance) to intentionally mislead or advantage individuals.

Some of the ways in which data can be distorted are:

* Changing the intervals; or scale on a graph could be inconsistent; or not starting at 0.
* No scale provided.
* Enhancing the visual representation of a graph, eg. 3D.
* The size of the bars on a bar graph may be enhanced.
* The sum of the sectors on a pie chart may not add to 100%.

**Purpose:**

You will learn to look critically at data presentations to avoid being misled.

**Illustrative example:**

The following graphs contain exactly the same data showing traffic accidents for the same years, but they look different. Critically analyse the graphs to show how they are misleading.

|  |  |
| --- | --- |
|  |  |

**Solution:**

The graphs seem to show that traffic accidents have increased substantially over the 4 years. This is due to the fact that the vertical axis does not start at zero. Furthermore, the graph on the left seem to show a steep increase because the scale on the vertical axis is not consistent from zero.

**Practice Exercise 9**

Refer to the following representations and answer the questions that follow.

1. Jenny’s History test scores are shown in the bar graph below:

|  |  |  |
| --- | --- | --- |
| **PERCENTAGE OBTAINED** | **70** | **Jenny’s History Test Scores** |
| **68** |  |  |  |  |  |  |  |  |  |
|  |  |  |
| **66** |  |  |  |  |  |  |  |  |  |  |
| **64** |  |  |  |  |  |  |  |  |  |
|  |  |  |
| **62** |  |  |  |  |  |  |  |  |  |
| **60** |  |  |  |  |  |  |  |  |  |
| **0** |  |  |  |  |  |  |  |  |  |  |
|  | **Test 1** | **Test 2** | **Test 3** | **Test 4** |  |
|  |  | HISTORY TEST |  |

Comment on the misleading test performance represented in the graph above.

1. The following two graphs represent the same information relating to a service providers sales of mobile devices, namely, tablets and cell phones. Study the graphs and then answer the question that follows:



 Give a reason why Graph 2 is misleading.

**Solution for Practice Exercise 9**

1. Inconsistent scale on the vertical axes creates the impression that the performance improved drastically. Incorrect perception.
2. The visual representation makes graph 2 look as though the ratio is more than double the bar as compared to the other bar.

**Summary Assessment – Data Handling**

1. Jerome achieved the following percentages in his five English tests:
60, 45; 30; 55 and 35.

If a pie chart is drawn to represent these marks, what angle would be represented by his last test result?

A. 96° B.12° C.56° D. 48°

 2. What type of data is represented by Joe’s weight of 57,5 kg?

 A. Discrete B. Continuous C.Categorical D. Qualitative

 3. Which measure of central tendency is affected if the 2 outliers from both ends of ordered data

 are removed?

1. Median B. Mode C. Mean D. None

 4. Virat Kholi, a famous Indian cricketer scored the following runs in 8 innings:

76; 58; 35; 40; 45; 0; 46; 50.

How many runs must he score in his next innings to maintain an average score of 50?

1. 50 B. 100 C. 200 D. 400

5. The box and whisker plot alongside shows the number of people visiting a hairdressing salon in the past 10 days.

 What percentage of people lie in the interval 14 to 18?

1. 15% B. 25% C. 50% D. 75%

6. The box and whisker plots below show the finishing times of girls and boys in an

 athletics meeting.

State whether the following statements are true or false giving reasons for each answer:

6.1 A girl was the first to cross the finishing line.

6.2 Boys had a higher median time.

6.3 Only a quarter of the boys completed the race when 50% of the girls completed the race.

6.4 The girls had the slowest finishing time.

6.5 The boys had a lower third quartile time.

 6.6 The girls had a higher interquartile range time than the boys.

 6.7 The boys performed better in the race overall.

 6.8 The data for the girls are negatively skewed.

7. The following graph was found in a local newspaper with the heading: “Inflation has gone down

rapidly in the last 4 years”. Do you agree with the newspaper? Motivate your answer.



 8**.** Refer to the following horizontal bar graphs showing 6 popular sports in S Africa according to

number of votes (in 1000’s), and then answer the questions that follow.

8.1 Write down the number of votes received by the most popular sport. State the sport.

 8.2 State which sport received the same number of votes. Write down the number of

 votes received.

8.3 Calculate the difference between the number of votes received for soccer and tennis.

8.4 Determine how many people were surveyed altogether.

**Solutions for Summary Assessment – Data Handling**

1. Total percentages = (60 + 45 + 30 + 55 + 35) = 225

Therefore last result angle calculation = (35 ÷ 225) × 360o = 56o

C.

 2. Since it’s a decimal, a part thereof it will be regarded as continuous.

 B.

 3. Median and mode won’t be affected but the mean will.

 C.

 4. There will be a total of 9 innings

Therefore to maintain an average of 50 the sum total will have to be:

50 × 9 = 450

The sum total before the 9th innings is: (76 + 58 + 35 + 40 + 45 + 0 + 46 + 50) = 350.

Therefore he has to score 450 – 350 = 100 in his 9th innings to maintain an average of 50.

B.

 5. This is quartile 1 to quartile 3: Therefore 50%.

 C.

 6.1 True: The lowest time recorded will be the fastest time. Girls had the lowest minimum time.

 6.2 True: Boys had a higher quartile 2 time.

 6.3 True: Boys quartile 1 = girls quartile 2.

 6.4 True: Girls had the highest maximum time.

 6.5 True: Boys third quartile was 38 while girls were 40.

 6.6 True: Girls IQR = 40 – 22 = 18 while boys IQR = 38 – 28 = 10

 6.7 True: Boys performance was consistent whereas girls’ performance was skewed to the right.

 6.8 False: Positively skewed.

 7. The vertical axis is not labelled, therefore no valid conclusion can be made.

 8.1 15 000 Soccer

 8.2 Netball and Cricket: 4 000

 8.3 15 000 – 3 000 = 12 000

 8.4 40 000

##### UNIT 5 – Probability

#### **Learning Outcomes**

By the end of the unit you should be able to:

* Express probability values in terms of common fractions, decimal fractions and

percentages.

* Use tree diagrams to determine the probability of dependent events.
* Effectively communicate conclusions and predictions that can be made from the

analysis and representation of data, using appropriate terminology such as,

trends, increase, decrease, constant, impossible, likely and even chance.

**Purpose:**

The purpose of probability is to find out the maximum percentage of an event occurring. Using probability you get to know about how likely an event will occur. In short, the reason we use probability is to make a prediction.

**Introduction**

* Probability is the **chance** of an event occurring or how likely something is to happen.
* An **experiment** is a situation involving probability resulting in outcomes.
* An **outcome** is the result of a single trial of an experiment.
* An **event** is one or more outcomes of an experiment.
* Probability of an event happening =

$$(Number of favourable (desired) outcomes)÷(Total outcomes)$$

* The chances of any event can be shown on a probability scale from 0 to 1:

a probability of 0 tells us that the event will never happen - it's impossible

a probability of 1 tells us that the event is certain to happen

a probability of ½ tells us that the event has an even chance of happening

* Probability scale below shows possible outcomes:
* Probability can be expressed as a fraction, decimal, percentage or using words (see scale above).
* For the purposes of this course dice is used to represent both singular (1 dice) and plural (more than 1 dice)

**Illustrative examples**:

1. Determine the probability of getting a number 4 using a single

six sided dice.

**Solution:**  How many numbers appear in a single six sided dice?

 This is called the total outcome.

 How many 4’s appear in a single six sided dice?

 This is called the desired outcome

Therefore, using the formula we get:

Probability = $\frac{Desired outcome}{Total outcome}$

 = $\frac{1}{6}$ as a fraction or 0,17 (as a decimal) or 17% (as a percentage)

1. Determine the probability of tossing a head on a coin.

Solution: How many outcomes are possible when you toss a

 single coin?

 How many heads are found on a single unbiased coin?

Therefore using the formula we get:

Probability = $\frac{Desired outcome}{Total outcome}$

 = $\frac{1}{2}$ (as a fraction) or 0,5 (as a decimal) or 50% (as a percentage)

1. In a bag there are 3 green balls, 5 red balls and 4 white balls.
	1. Determine the probability of randomly selecting a white ball.

**Solution**: What is the total number of balls altogether?

 How many white balls there are?

Therefore the probability = $\frac{4}{12}$ = $\frac{1}{3}$ or 0,3 or 33%.

* 1. Determine the probability of randomly selecting a blue ball.

**Solution**: Impossible as there are no blue balls; or 0

* 1. John selected a ball from the bag and it was green. He did not replace the ball back into the bag. He now wants to select another green ball.

Determine the probability of John now randomly selecting a green ball from the bag.

**Solution**: How many balls are now left in the bag?

How many green balls are now in the bag?

Therefore the probability = $\frac{Number of green balls left in the bag}{Total number of balls in the bag}$

 =$\frac{2}{11}$ or 0,18 or 18%

**Activity for guided reflection**

**Practice Exercise 9**

Determine the following probabilities expressing all your answers as a fraction, decimal and a percentage.

**Guided reflections:**

**You need to establish what the total outcome for an event is.**

**You also need to know what the desired outcome is.**

**In cases where there is a second selection you need to know:**

* **Whether there was a replacement – replace the item thereby not interfering with the total outcome**
* **Or whether the item was not replaced – in this case the total outcome will be reduced by one.**

Determine the probability of randomly:-

* 1. Selecting an even number on a six sided dice.
	2. Choosing a white Smarties from a box containing 8 red, 12 blue and 6 green Smarties.
	3. Selecting a second Smarties from Question 1.2 and it has to be red, after having already eaten one blue smarties.
	4. Selecting a picture card from a pack of 52 playing cards. Note: Jack, Queen and King in each suite of cards are considered picture cards. There are 4 suites of cards in each pack of playing cards.
	5. Driving a silver car in an auto dealership which has the following cars on the floor: 10 white; 3 silver; 5 blue and 2 red cars.

**Solution – Practice exercise 9**

* 1. Even numbers found on a six side dice: 2; 4; 6

Therefore the probability = $\frac{3}{6}=\frac{1}{2}$ or 0,5 or 50%

 1.2 0 or impossible or 0%

 1.3 After eating one smarties we now have 25 left

 Therefore probability = $\frac{8}{25}$ or 0,32 or 32%

 1.4 52 playing cards: 4 suites each having 3 picture cards = 12 picture cards altogether

 Therefore probability = $\frac{12}{52}=\frac{3}{13}$ or 0,23 or 23

 1.5 Probability = $\frac{3}{20}$ or 0,15 or 15%

## Tree Diagrams

A tree diagram is a visual graphical method of obtaining the outcomes of a random experiment.

**Purpose**

Tree diagrams allow you to see all the possible outcomes of an event and calculate their probability. Each branch in a tree diagram represents a possible outcome.

**Introduction**

* A tree diagram is a means of representing a sequence of events.
* Tree diagrams enable us to see all the possible outcomes of an event and calculate their probability.
* Each branch in a tree diagram represents a possible outcome.
* For each pair of branches the sum of the probabilities adds to 1.
* "And" only means multiply outcomes if events are independent, that is, the outcome of one event does not affect the outcome of another.
* "Or" means add the outcomes.
* Remember we often use the word “unbiased” or “fair” to indicate that we are dealing with for example a coin that has a head and a tail and not two heads or two tails.

**Illustrative example**

A fair coin is tossed twice. Determine the probability of getting two tails (TT)

Solution:

Study the tree diagram: There are two branches for the first toss (Heads and tails)

 The probability of each branch is written on the branch (0,5 or $\frac{1}{2}$ )

 The outcome is written at the end of the branch, namely Head (H) or Tail (T)

We now extend the tree diagram to include the second toss.

 The probability once again is written on each branch.

The probability for each outcome is obtained by multiplying the probability along the branches at the end, namely TWO HEADS (HH) = 0,5 X 0,5 = 0,25 or $\frac{1}{4}$

Therefore probability of getting two tails: P(TT) =0,25 or $\frac{1}{4}$



**Practice Exercise 10**

1. Moondowns Football Club (FC) played two matches in May. There are THREE possible outcomes for each match: W (win), L (lose) or D (draw).

Study the following tree diagram and then answer the questions which follow

(Adapted from DBE, November, 2008, Paper 2)



* 1. Complete the tree diagram to show ALL the possible outcomes of the two matches.
	2. Use the tree diagram to predict the probability that Moondowns FC will:
1. Win both matches
2. Win only one of the matches
3. Draw at least one of the matches.
4. A bag contains the following coloured marbles: 5 red; 4 white and 3 green. Study the tree diagram that follows and then answer the questions which follow:



* 1. Sizwe picks a marble from the bag without looking but then places it back into the bag. He then picks another marble for the second time. Determine the probability of choosing at least one green marble in his two picks.
	2. Sizwe picks a marble and then places it back. He then picks another marble. What is the probability that he chooses no white marbles in his two picks?
1. A coin and a six sided dice are thrown at random. Find the probability of:
	1. Getting a Tail and an Even number.
	2. Getting a Head and a prime number.



**Solutions – Practice exercise 10**

* 1. 

 1.2 (a) Win both matches = $\frac{1}{9}$ or 0,11 or 11,11% or $\left(\frac{1}{3}×\frac{1}{3}=\frac{1}{9}\right)$ Multiplying probability

 along the path of the two branches leading to WW.

1. Win only one of the matches

 Four of the outcomes have only one W

 Therefore probability = $\frac{4}{9}$ or 0,44 or 44,44% or multiplying along the path of the two

 branches and adding them all up as follows: $\left(\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}=\frac{4}{9}\right)$

1. Five of the outcomes have at least one D

 Therefore the probability = $\frac{5}{9}$ or 0,56 or 55,56% or multiplying along the path of the two

 branches and adding all 5 outcomes as follows: $\left(\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}=\frac{5}{9}\right)$

2.1 $\frac{5}{9}$ as five out of the nine outcomes have at least one green.

* 1. $\frac{4}{9}$ as there are 4 outcomes that have no “W” in the final outcome.
	2. $\frac{3}{12}=\frac{1}{4}$
	3. $\frac{3}{12}=\frac{1}{4}$

**Guided reflection**

So there you go, when in doubt draw a tree diagram, multiply along the branches and add the columns. Make sure all probabilities add to 1 and you are on your way to understanding probability type questions..

**Summary Exercise**

1. Coloured marbles are placed in three different cups as follows:

 First cup: 2 Blue marbles

 Second cup: 2 Red marbles

 Third cup: 1 red and 1 blue marble

Choose one word from the following that best describes each of the statements that follow:

 Certain; Impossible; Likely

* 1. Choosing a blue marble from the first cup.
	2. Choosing a blue marble from the second cup.
	3. Choosing a blue marble from the third cup.
1. Sarah’s shopping bag contains the following:

7 red peppers, 2 green peppers, 3 purple peppers and 4 yellow peppers.

Without looking Sarah takes out a pepper and eats it.

* 1. Write down the probability it is not a red pepper.
	2. Sarah realizes she had eaten a green pepper which tasted good. She digs into her shopping bag without looking for a second green pepper.

Determine the probability it will be a green pepper she chooses.

1. The table below shows the number of left-handed and right handed cricket players in a school.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Left-handed** | **Right-handed** | **Total** |
| **Male** | **2** | **18** | **20** |
| **Female** | **5** | **25** | **30** |
| **Total** | **7** | **43** | **50** |

A cricket player is chosen at random:

* 1. Determine the probability the player is left-handed.
	2. Determine the probability that the player is male and right handed.
1. The following bar graph shows the number of correct answers per question in a test which had 5 questions.

 Determine the probability that a question chosen at random from this data:

* 1. has at least 3 correct answers
	2. has less than 5 correct answers
	3. has 2 or more correct answers.

1. Two six sided dice are thrown together. Use a tree diagram to find the probability that one number is odd and the other number is even.

**Solutions – Summary exercise**

* 1. Probability of choosing a blue from cup 1 is 100% as there are only blues in cup 1

Therefore “Certain”.

* 1. Probability of choosing a blue marble from cup 2 is 0 as there are no blue marbles in cup 2

Therefore “Impossible”.

* 1. Probability of choosing a blue marble from cup 3 is 50% as there is one blue and one red

Therefore “Likely”.

* 1. There are 16 peppers altogether. Of these 7 are red which means the rest (9) are not red

 Therefore the probability of choosing a pepper that is not red = $\frac{9}{16}$

2.2 Since a green pepper was eaten, there would be one less green implying one less in the

 Total. Therefore the probability = $\frac{2-1}{16-1}=\frac{1}{15}$

* 1. Probability of left-handed cricketer = $\frac{7}{50}$
	2. Probability of male and right-handed = $\frac{18}{50}$=$\frac{9}{25}$

4.1 $\frac{11}{15}$

4.2 $\frac{8}{15}$

4.3 $\frac{13}{15}$

5. The use of the word “and” means add.

 There are six possible scores on one die: {1, 2, 3, 4, 5, 6}
 Of these, three are even: {2, 4, 6} and three are odd: {1, 3, 5}

 Therefore, the tree diagram looks as follows:



 So the probability that one number is even and the other is odd = $\frac{1}{4}$ + $\frac{1}{4}=\frac{1}{2}$