NCCA
An Chomhairle Náisiúnta Curaclaim agus Measúnachta
National Council for Curriculum and Assessment

Project Maths
Tionscadal Mata

Project Maths

Mathematics Resources for Students

Leaving Certificate – Strand 1

Statistics and Probability

INTRODUCTION

This material is designed to supplement the work you do in class and is intended to be kept in an A4 folder. Activities are included to help you gain an understanding of the mathematical concepts and these are followed by questions that assess your understanding of those concepts. While there are spaces provided in some activities/questions for you to complete your work, you will also need to use your copybook/A4 pad or graph paper. Remember to organise your folder so that it will be useful to you when you revise for tests and examinations. As you add pages to your folder, you might consider dating or coding them in a way that associates them with the different topics or syllabus sections. Organising your work in this way will help you become personally effective. Being personally effective is one of the five key skills identified by the NCCA as central to learning (www.ncca.ie/keyskills). These key skills are important for all students to achieve their full potential, both during their time in school and into the future.

As you work through the material in this booklet and with your teacher in class, you will be given opportunities to develop the other key skills. You will frequently work in pairs or groups, which involves organising your time effectively and communicating your ideas to the group or class. You will justify your solutions to problems and develop your critical and creative skills as you solve those problems. As you complete the activities you will be required to process and interpret information presented in a variety of ways. You will be expected to apply the knowledge gained to draw conclusions and make decisions based on your analysis. The sequence in which the sections/topics are presented here is not significant. You may be studying these in a different order, or dipping in and out of various sections over the course of your study and/or revision.

The questions included in this booklet provide you with plenty of opportunities to develop communication skills and to promote mathematical discourse. When your teachers mark your work they will gain insights into your learning and will be able to advise you on what you need to do next.

The material in the booklet is suitable for both Junior Certificate and Leaving Certificate, since at Leaving Certificate you consolidate and build on the concepts you learned at Junior Certificate and, occasionally, you may need to explore some of the activities or exercises which were used to introduce these concepts. Through completing the activities and questions contained in this booklet, you will develop a set of tools that will help you become a more effective learner and these tools can be used across the curriculum. Solving problems of this nature should also improve your confidence in doing mathematics, thus helping you to develop a positive attitude towards mathematics and to appreciate its role in your life.

The mathematics syllabus documents can be accessed at www.ncca.ie and you will find other relevant material on www.projectmaths.ie.

PROBABILITY 1

SYLLABUS TOPIC: CONCEPTS OF PROBABILITY

LEARNING OUTCOME

As a result of completing the activities in this section you will be able to

- decide whether an everyday event is likely or unlikely to happen
- recognise that probability is a measure on a scale of 0 - 1 of how likely an event is to occur.
- connect with set theory; discuss experiments, outcomes, sample spaces
- use the language of probability to discuss events, including those with equally likely outcomes

INTRODUCTION

The activities described below and the questions that follow give you the opportunity to reinforce your understanding of the basic concepts of probability. The activities are designed to build on previous experiences where you estimated the likelihood of an event occurring. Some of the activities will be done in class under the direction of your teacher; others can be done at home.

Activity 1.1

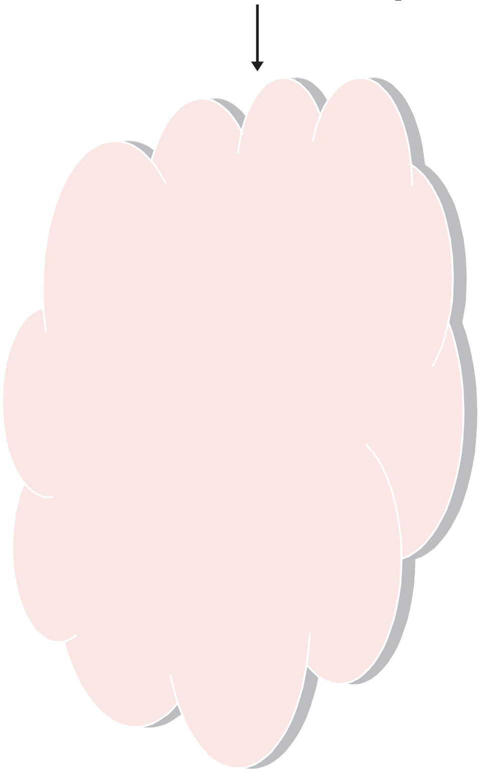
A probability describes mathematically how likely it is that something will happen. We can talk about the probability it will rain tomorrow or the probability that Ireland will win the World Cup.

Consider the probability of the following events

- It will snow on St Patrick's day
- It will rain tomorrow
- Munster will win the Heineken Cup
- It is your teacher's birthday tomorrow
- You will obtain a 7 when rolling a die
- You will eat something later today
- It will get dark later today

Words you may decide to use: certain, impossible, likely, very likely

Student Activity

Certain not to happen	Area of Uncertainty	Certain to happen
1.		1.
2.		2.
3.		3.

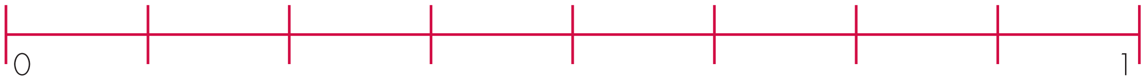
Phrases used to describe uncertainty

1.
2.
3.
4.
5.

Use the table provided or mark your work page out in a similar way and place each of the events in the approriate section.Note the phrases you used to describe uncertainty.

Activity 1.2

The Probability Scale



Extremely unlikely	50/50	3/8	1 in 4 chance
Probability of getting an odd number when rolling a die	87.5%	Extremely likely	1/2
	0.125	3/4	Impossible
1/4	Certain	75%	1
Equally likely	0.25	0	

1. Place the above phrases, numbers and percentages at the correct position on the probability scale.
2. Find and write down instances from TV, radio, or in the newspaper which illustrate how probability affects people's lives.

Questions

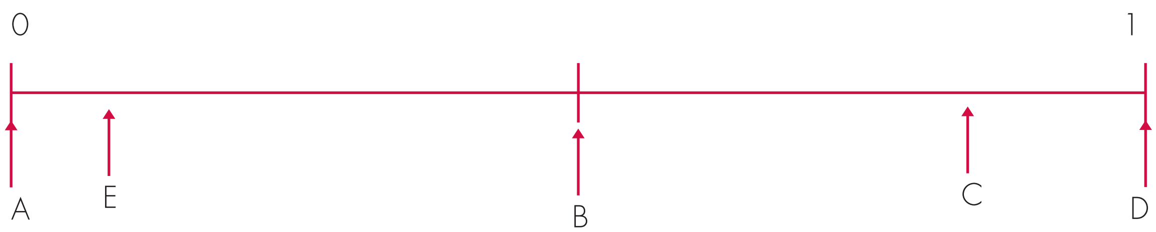
Q.1 For each event below, estimate the probability that it will happen and mark this on a probability scale.

- It will snow in Ireland on August 16th
- Your maths teacher will give you homework this week
- Your will eat fish later today
- You will go to bed before midnight tonight
- You will go to school tomorrow

Q. 2 Use one of the words certain, likely, unlikely, impossible to describe each of the events below. Give a reason for each of your answers.

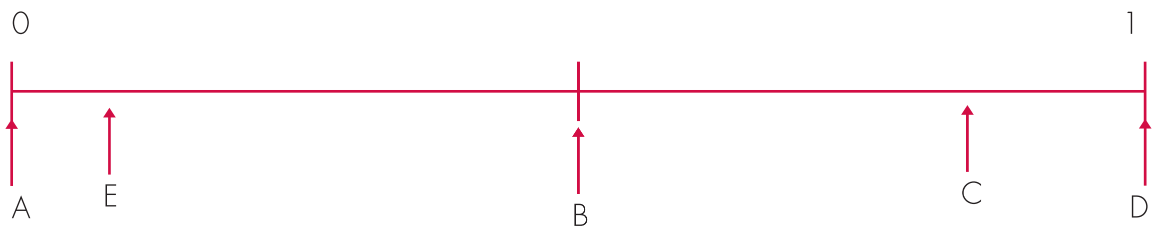
- You are more than 4 years old
- You will arrive on time to school tomorrow
- You will miss the school bus tomorrow
- Your county will win the Championship this year.

Q. 3 The probability line shows the probability of 5 events A, B, C, D and E



a. Which event is certain to occur?
b. Which event is unlikely but possible to occur?
c. Which event is impossible?
d. Which event is likely but not certain to occur?
e. Which event has a 50:50 chance of occurring?

Q. 4 The events A, B, C, D have probabilities as shown on this probability line;



i. Which event is the **most likely** to take place?

ii. Which event is the **most unlikely** to take place?

iii. Which event is **more likely than not** to take place?

Q. 5 When you toss an unbiased coin the probability of getting a head is ½, because you have an equal (or even) chance of getting a head or tail. Name two other events that have a probability of ½.

Q. 6 The 'events' A, B, C, D are listed below;

A: You will live to be 70 years old
B: You will live to be 80 years old
C: You will live to be 100 years old
D: You will live to be 110 years old

Make an estimate of the probability of each event, and place it on a probability scale.

Q. 7 Sarah and Alex are exploring probability and Sarah has these cards:

2	3	9	5	6	7	8	9
△	△	■	△	■	■	△	△

Alex takes a card without looking. Sarah says

On Alex's card ■
is more likely than △

i. Explain why Sarah is wrong.

ii. Here are some words and phrases that can be associated with probability:

impossible	not likely	certain	likely

Choose a word or a phrase to fill in the gaps below.

It is .... that the number on Alex's card will be smaller than 10.

It is .... that the number on Alex's card will be an odd number.

Sarah mixes up the cards and places them face down on the table.
Then she turns the first card over, like this:

Alex is going to turn the next card over

iii. Complete the sentence:

On the next card, __________ is less likely than __________

The number on the next card could be higher than 5 or lower than 5

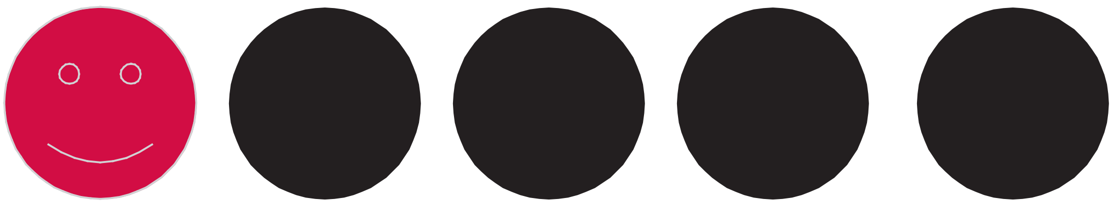
iv. Which is more likely? Tick the correct box below.

☐ Higher than 5 ☐ Lower than 5 ☐ Cannot tell

Explain your answer.

Q. 8 Lisa has some black counters and some red counters.

The counters are all the same size.
She puts 4 black counters and 1 red counter in a bag.



a. Lisa is going to take one counter out of the bag without looking.

She says:

> There are two colours, so it is just as likely that I will get a black counter as a red counter.

i. Explain why Lisa is wrong. What is the probability that the counter she takes out is black?

ii. How many more red counters should Lisa put in the bag to make it just as likely that she will get a black counter as a red counter?

b. Jack has a different bag with 8 counters in it. It is more likely that Jack will take a black counter than a red counter from his bag.

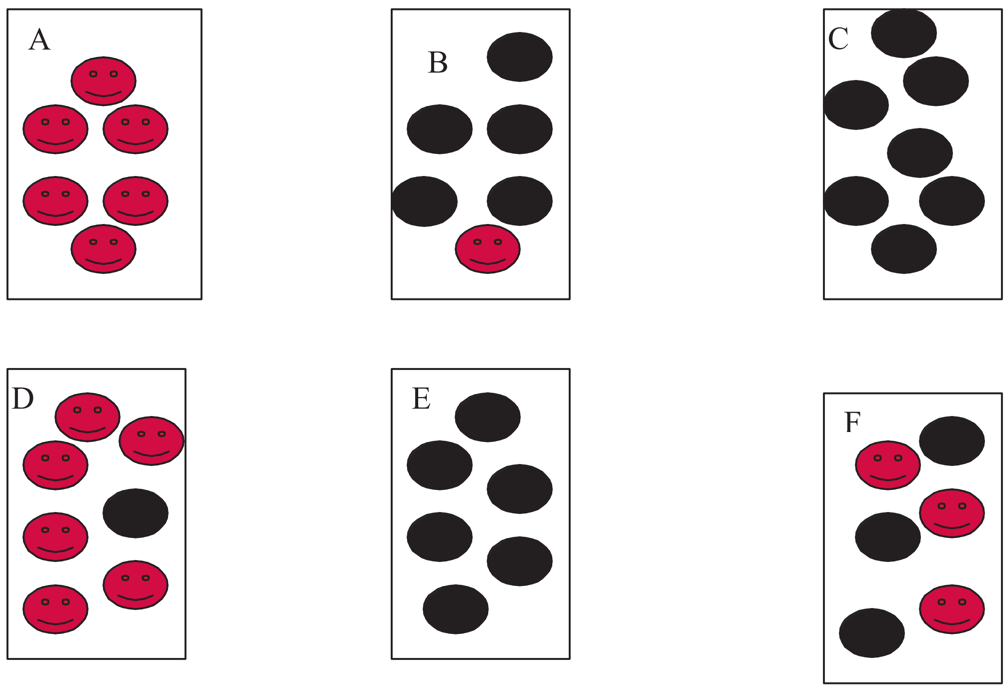
iii. How many black counters might there be in Jack's bag? Suggest a number and explain why this is a possible answer.

c. Jack wants the probability of taking a black counter from his bag to be the same as the probability Lisa had at the start of taking a black counter from her bag, so he needs to put extra counters into his bag.

iv. Assuming Jack had the number of black counters you have suggested at (iii) above, how many extra black counters and how many extra red counters (if necessary) should Jack put in his bag?

Explain your reasoning.

Q. 9 (a) Josh has some boxes containing red and black counters.



He is going to take a counter from each box without looking.

a. Match boxes (using the letters A-F) to the statements below. Explain your reasoning each time.

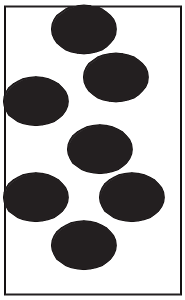
It is **impossible** that Josh will take a black counter from box........because

It is **equally likely** that Josh will take a black or red counter from box........because

It is **likely** that Josh will take a red counter from box............because

It is **certain** that Josh will take a black counter from box...........because

Josh selects box C which has 7 black counters in it



He wants to make it **more likely** that he will take a red counter than a black counter out of the box.

How many red counters must he put into the box? Explain your answer.

b. In another box, there are 30 counters which are either red or black in colour.
It is **equally likely** that Josh will take a red counter or a black counter from the box.
How many red counters and how many black counters are there in the box?

c. Extension question
There are 40 counters in a box which are either red or black in colour.
There is a **75% chance** that Josh will take a red counter from the box.
How many black counters are in the box? Explain your answer.

PROBABILITY 2

SYLLABUS TOPIC: CONCEPTS OF PROBABILITY

LEARNING OUTCOME

As a result of completing the activities in this section you will be able to

- estimate probabilities from experimental data; appreciate that if an experiment is repeated, there will be different outcomes and that increasing the number of times an experiment is repeated generally leads to better estimates of probability
- associate the probability of an event with its long run relative frequency

INTRODUCTION

The activities described below and the questions that follow give you the opportunity to reinforce your understanding of the basic concepts of probability. You begin by rolling two coins and progress to playing a game involving rolling two dice. You will use a sample space to list all the possible outcomes and begin to consider the concept of expected value as you investigate the idea of fairness in relation to the game.

Activity 2.1

Toss two coins simultaneously about 30 times and record all the outcomes.
Do you notice any outcomes coming up over and over again?
Do some of these come up more frequently than others?
Use the grid below to show the 4 possible outcomes (the sample space) of heads (H) and tails (T).

		Coin 1	
		H	**T**
Coin 2	**H**		
	T		

Use the sample space to calculate the probability of each outcome occurring (i.e. the theoretical probability).
From the results you obtained in the 30 tosses, construct a table showing the number of times each outcome occurred and its relative frequency. Compare these to the theoretical probability.

Outcome	Tally	Relative Frequency

Activity 2.2

Working in pairs, roll a die 30 times (i.e. 30 trials) and enter your results into a table similar to the one outlined below

Number which appears on die (outcome of trial)	**How many times did this happen?** (Use tally marks to help you count.)	**Total** (frequency)
1		
2		

As you complete your own table compare it with that of another group.
Are there any similarities?

Your teacher may ask you to complete a Master sheet showing the results of all the groups in the class (a total of N trials).

Outcome of trial	**Frequency (group results)**	**Total of frequencies**	**Relative frequency)** $\frac{\textit{Total of frequencies}}{\textit{sample size (N)}}$	**% of total scores Rel. Freq × 100**	**Probability**
1	E.g. 5+6+5+...				
2					
3					
4					
5					
6					
		SUM			

The sum of all the relative frequencies is

The sum of all the percentages is

The sum of all the probabilities is

Conclusion:

What does your experiment tell you about the chance or probability of getting each number on the die you used?

Your die can be described as being unbiased. Can you explain why?

Activity 2.3

a. Each student tosses a coin 30 times and records their results for every 10 tosses.

No of tosses	No of Heads	Relative frequency
10		
10		
10		

b. What does the table you completed in (a) tell you about the probability of getting a head?

c. Now put all the results for the class together and obtain a new estimate of the probability of getting a head.

d. Is your new estimate closer to ½ than the estimate in (a)?

Activity 2.4

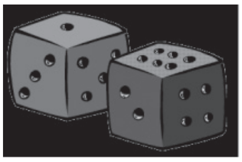
This is a game for two players, A and B. They take turns to roll two dice and add the two numbers shown on each toss. The winner is determined as follows:

A wins if the sum of the numbers on the dice (i.e. outcome) is 2, 3, 4, 10, 11 or 12.

B wins if the sum of the numbers on the dice is 5, 6, 7, 8, 9.

Before you begin predict which player is most likely to win.

I think player ______ will win because ______



Play the game until one player reaches the bottom of the game sheet.

GAME SHEET

	A wins	**A wins**	**A wins**	B wins	B wins	B wins	B wins	B wins	**A wins**	**A wins**	**A wins**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**

Record the number of times each player wins in the table below. The relative frequency is the **total no. of wins divided by the total no. of games.**

	Total (frequency)	**Relative frequency**
Player A wins		
Player B wins		
Totals		

As a class exercise construct a Master Tally sheet and record the results of the whole class

	Total (frequency)	**Relative frequency**
Player A wins		
Player B wins		
Totals		

Does your predicted result agree with your actual result? Think about why this happens.
Complete the table below showing all the possible outcomes for throwing two dice.

	1	**2**	**3**	**4**	**5**	**6**
1	(1,1)					
2						
3			(3,4)			
4					(4,6)	
5						
6						

In the case of equally likely outcomes, the probability is given by the number of outcomes of interest divided by the total number of outcomes.

Construct a table to show the probability of each outcome above, with the probability = $\dfrac{\text{no of outcomes in the event}}{\text{no of outcomes in the sample space}}$

Sum of two dice	Frequency	Probability
2	1	1/36
3	2	2/36

Look back at the rules of the game.

Original Rules: Player A wins when the sum is 2, 3, 4, 10, 11 or 12.

Player B wins when the sum is 5, 6, 7, 8 or 9.

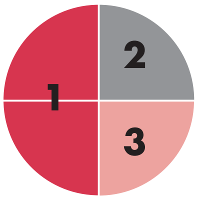
For how many outcomes will player A win?_______

For how many outcomes will player B win?_______

Does the game seem fair? If not, suggest a change to the rules which would make it fairer.

Create a mind map or a graphic organiser (http://www.action.ncca.ie) that will help you remember how to calculate the relative frequency of an event occurring.

Q. 1 Sophie and Andrew are playing a game with a fair, six-sided die and the spinner shown. They throw the die and spin the spinner simultaneously and note the total



Sophie I will carry your bag home if the total is **2, 3, 8** or **9**. You carry mine if the total is **4, 5, 6** or **7**

Andrew said

Create a sample space showing the possible outcomes and use it to help Sophie decide whether or not she should play the game. Justify your advice to Sophie.

Q. 2 What is the probability of getting a head and a 6 when you simultaneously toss a fair coin and roll a fair, six-sided die?

How would this probability change if the die was replaced with:

a. A four-segment spinner (segments of equal area) numbered 1, 6, 6, 5?

or

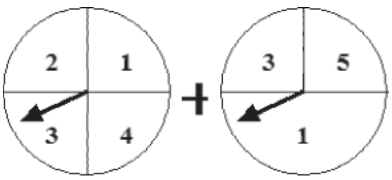
b. A suit of spades from a deck of playing cards (and 1 card is chosen at random from the suit)?

Q. 3 A spinner has four unequal sections, red, black, pink and grey.

The probability that the spinner will land on red is 0.1 [P(red) =0.1]
The probability that the spinner will land on black is 0.2 [P(black) =0.2]
The probability that the spinner will land on pink is the same as the probability that it will land on grey.

Calculate the probability that the spinner will land on grey. Justify your answer.

Q. 4 Design a spinner that simulates in 1 spin the sum of the outcome from spinning the two spinners below. Explain and justify your design.



Q. 5 A calculator can be used to generate random digits. Sandra generates 100 random digits with her calculator. She lists the results in the table below.

0	卌 ||	**5**	卌 卌
1	卌 |||	**6**	卌 卌 ||||
2	卌 |	**7**	卌 卌 ||
3	卌 卌 ||	**8**	卌 卌 |
4	卌 卌 ||	**9**	卌 卌 ||||

Based on Sandra's results, estimate the probability that the calculator produces:
a) 9, b) 2, c) a digit that is a multiple of 3, d) a digit that is prime.

Q. 6 Four students each threw 3 fair dice.

They recorded the results in the table below.

Name	**Number of throws**	**All different numbers**	**Exactly 2 numbers the same**	**All 3 numbers the same**
Jane	50	36	12	2
Paul	150	92	45	13
Tom	40	18	20	2
Patti	120	64	52	4

a. Which student's data are **most likely** to give the best estimate of the probability of getting

All numbers the same ☐ Exactly 2 numbers the same ☐ All 3 numbers the same ☐

Explain your answer.

b. This table shows the students' results collected together:

Number of throws	All different	Exactly 2 numbers the same	All 3 numbers the same
360	210	129	21

Use these data to estimate the **probability** of throwing numbers that are **all different**.

c. The theoretical probability of each result is shown below:

	All Different	2 the same	All the same
Probability	5/9	5/12	1/36

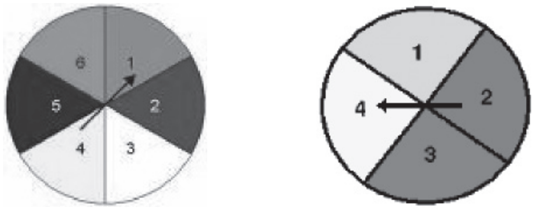
Use these probabilities to calculate, for 360 throws, **how many times** you would theoretically expect to get each result. Complete the table below.

Number of throws	All different	2 the same	All the same
360			

d. Give a reason why the students' results are not the same as the theoretical results.

Think: How would this question be different if coins, spinners or playing cards were used?

Q. 7 Pierce and Bernie were investigating results obtained with the pair of spinners shown.



They used a table to record the total of the two spinners for 240 trials. Their results are given in one of the three tables A, B and C below.

Table A

Sum	Frequency	Relative frequency
2	10	1/24
3	20	1/12
4	30	1/8
5	30	1/8
6	60	1/4
7	40	1/6
8	20	1/12
9	20	1/12
10	10	1/24
Total	240	1

Table B

Sum	Frequency	Relative frequency
2	12	12/240
3	12	12/240
4	27	27/240
5	27	27/240
6	35	35/240
7	45	45/240
8	24	24/240
9	18	18/240
10	40	40/240
Total	240	

Table C

Sum	Frequency	Relative frequency
2	11	
3	19	
4	32	
5	30	
6	29	
7	28	
8	17	
9	14	
10	60	
Total	240	

Complete the relative frequency column in table C.
Use your results to decide which, if any, of these three tables might represent the results found by Pierce and Bernie. Explain your reasoning.

Q. 8 A spinner with 3 equal segments numbered 1, 2 and 3 is spun once.

i. Give the sample space of this experiment.

ii. What is the probability that the spinner stops on number 2?

iii. What is the probability that the spinner stops on a number greater than or equal to 2?

Q. 9 Pierce and Bernie were investigating the results given by the spinner shown, by spinning it 60 times and recording the results.

Their results are given in one of the three tables below, A, B and C

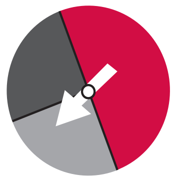


Table A

result	tally	count
red	𝍸 𝍸 𝍸 𝍸 𝍸 I	21
grey	𝍸 𝍸 𝍸 IIII	19
black	𝍸 𝍸 𝍸 𝍸	20

Table B

result	tally	count
red	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 II	47
grey	𝍸 I	6
black	𝍸 II	7

Table C

result	tally	count
red	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 II	32
grey	𝍸 𝍸 𝍸	15
black	𝍸 𝍸 III	13

a. Which of the three tables above is most likely to be like the one that Pierce and Bernie made? Explain how you made your decision.

b. For each of the other two tables, draw a diagram of a spinner that is likely to produce results like those shown in each table.

The following questions represent an application of the concept of probability. They are most suited for LCHL. They require you to have an understanding of the fact that probability is a quantity that gives a measure on a scale of 0-1 (0-100%) of how likely an event is to occur. They also require you to display an understanding that, if an experiment is repeated, there will be different outcomes and increasing the number of times an experiment is repeated generally leads to better estimates of probability. Discussing such probability scenarios with your peers in class allows you to make sense of probability and its application to real life phenomena. When explaining your answers you are displaying your competency in the key skill of communication.

Q. 10 Ten patients with HIV are each given a new treatment which the specialist says has a 30% chance of completely curing them.

What can each patient deduce from this?

Q. 11 The weather forecast for tomorrow states there is a 25% chance of rain in Leinster.

What exactly does this statement mean?

PROBABILITY 3

SYLLABUS TOPIC: CONCEPTS OF PROBABILITY

LEARNING OUTCOME

As a result of completing the activities in this section you will be able to

- apply the principle that in the case of equally likely outcomes the probability is given by the number of outcomes of interest divided by the total number of outcomes
- use binary/counting methods to solve problems involving successive random events where only two possible outcomes apply to each event
- discuss basic rules of probability (AND/ OR, mutually exclusive) through the use of Venn Diagrams
- find the probability of intersection of two independent events

HL learners will

- extend your understanding of the basic rules of probability (AND/ OR, mutually exclusive) through the use of formula (addition rule):

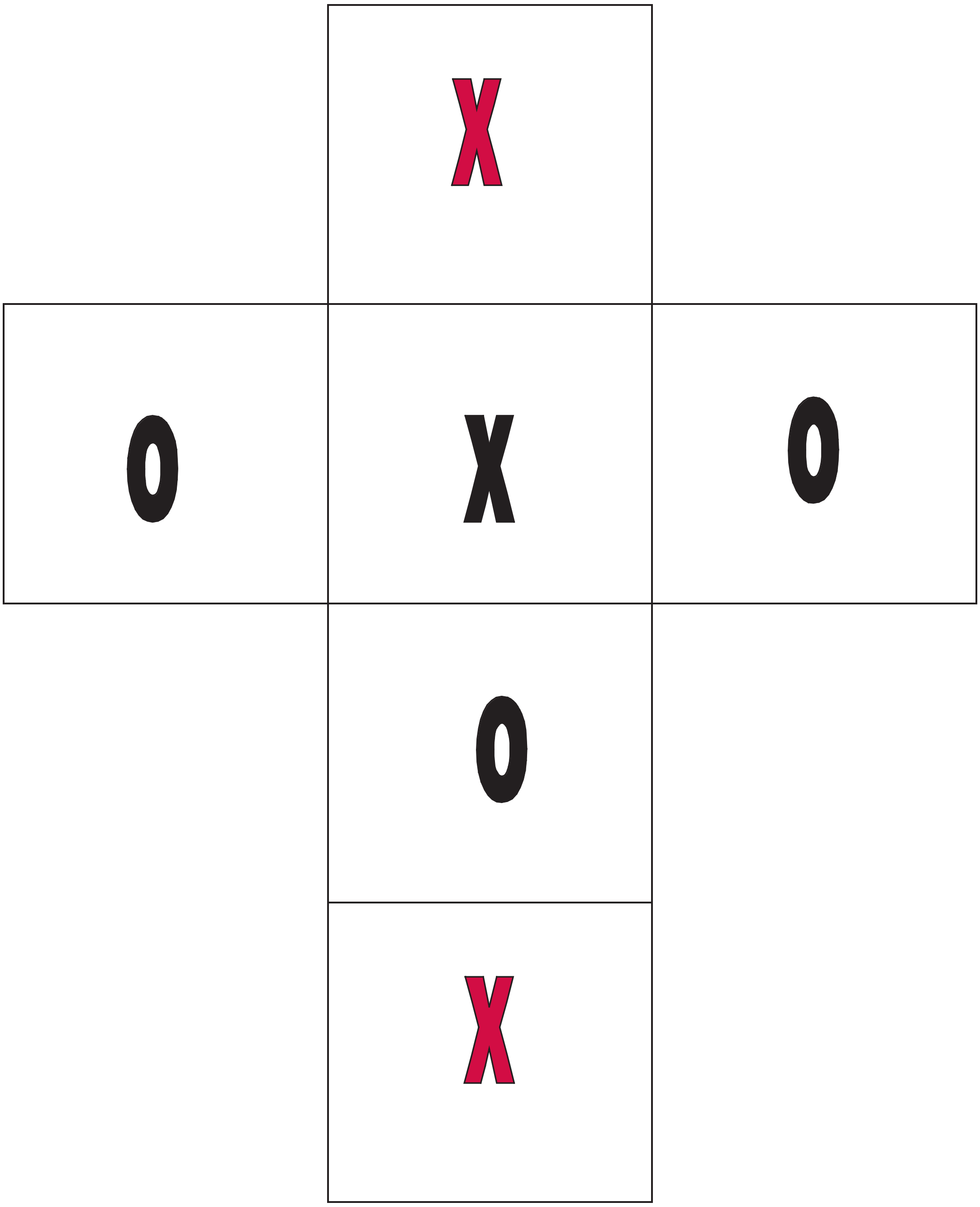
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

INTRODUCTION

The activities described below are intended for JC HL, LC OL and LC HL students.
The questions that follow the activities allow you to construct an understanding of outcome spaces and of the 'and' and 'or' rules of probability and use them to solve and compose problems.

Activity 3.1

Fold the net shown overleaf to form a cube.
Roll the cube; there are 6 possible outcomes.
Complete the boxes to show the 6 possible outcomes



a.

i. What is the probability of getting a cross?

ii. What is the probability of getting a circle?

iii. What is the probability that you will get a red symbol?

iv. What is the probability that you will get a black symbol?

Explain how you arrived at your answer.

b. Let's take the probability of getting a cross.

i. What will happen to this probability if I say the cross has to be black?

ii. What will happen to this probability if I say the cross has to be red?

How do these answers relate to your answer in (a) (i)?

c. Now look at the probability of getting a circle.

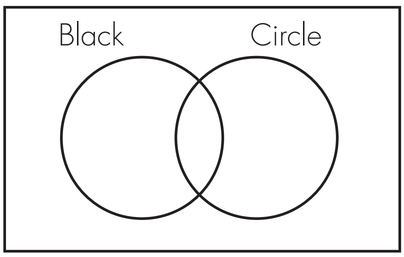
i. What will happen to the probability if I say the circle has to be red?

ii. What will happen to the probability if I say the circle has to be black?

If two events cannot possibly happen together we say they are ***Mutually Exclusive***

d. What events are mutually exclusive in the example above?

Represent this situation in the Venn diagram



We can write the questions in mathematical language

P (Closed ∩ Cross) means the probability of the symbol being closed AND a cross

Write mathematical sentences for the following and give the answer in each case

i. The probability of the symbol being black AND a circle
ii. The probability of the symbol being red AND a circle
iii. The probability of the symbol being red AND a cross

e. What about OR?

Think about this question: What is the probability of the symbol being black OR a circle? Can you see this in the Venn diagram above?

[HL] Use the Venn diagram to help you write an equation for P(Black ∩ Circle) in terms of the P(Black) and P(Circle)

f. Now think about the probability of the symbol being black OR a cross P(Black U Cross)

Can you see this in the Venn diagram?

HL Use the Venn diagram to help you write an equation for P (Black U Cross) in terms of the P(Black) and P(Cross)

Activity 3.2

a. Suppose you were asked to randomly choose one number from the numbers 1 to 10.

i. What is the probability that the number you choose would be greater than 5?
ii. What is the probability that the number you choose would be even?
iii. What is the probability that the number you choose would be greater than 5 AND even?
iv. What is the probability that the number you choose would be greater than 5 OR even?

b. Write the numbers 1 to 10 on ten pieces of paper. Now fold the pieces of paper so that you can no longer see the numbers on each piece. Working in pairs, place the pieces of paper in a hat or a bag and ask a classmate to choose a piece of paper from the hat/bag. Get your classmate to record the number they chose in a table similar to this.

Number	1	2	3	4	5	6	7	8	9	10
Frequency										

Place the piece of paper back into the bag/hat and stir the pieces of paper around. Now repeat this experiment until you have done it 20 times and combine the results from the other pairs in the class.

i. How many times was a number greater than 5 chosen?

ii. How many times was an even number chosen?

iii. How many times was the number chosen greater than 5 AND even.

iv. How many times was the number chosen greater than 5 OR even.

Use these answers to compute relative frequencies for each of the events listed above and compare these relative frequencies with the probabilities you obtained earlier. Are they the same?

You may like to look back at **Probability 2** to remind yourself how relative frequencies are calculated.

What do you expect would happen to the relative frequencies if you were to repeat this experiment one million times?

The following questions provide you with the opportunity to

- connect with JC set theory to discuss experiments, outcomes, sample spaces
- discuss basic rules of probability (AND/ OR, mutually exclusive) through the use of Venn diagrams

You will need to recall your work on the binary operations of addition and subtraction of fractions

Q. 1 In a class 1/2 of the pupils represent the school at winter sports and 1/3 represent the school at summer sports and 1/10 at both. Draw a Venn diagram to represent this. If a pupil is chosen at random, what is the probability that someone who represents the school at sport will be selected?

Q. 2 In a certain street 1/5 of the houses have no newspaper delivered, 1/2 have a national paper delivered and 1/3 have a local paper delivered. Draw a Venn diagram to represent this information and use it to find the probability that a house chosen at random has both papers delivered.

Q. 3 Consider the following possible events when two dice, one red and one green, are rolled.

A: The total is 3
B: The red is a multiple of 2
C: The total is ≤ 9
D: The red is a multiple of 3
E: The total is ≥ 11
F: The total is ≥ 10

a. Which of the following pairs are mutually exclusive? Discuss your choice with a classmate.

(i) A,D (ii) C, E (iii) A ,B (iv) C,F (v) B,D (vi) A,E

b. Write a definition explaining what is meant by saying that two events are mutually exclusive.

Q. 4 The probability that Josh will be in the school football team is 0.6, the probability that he will be in the hurling team is 0.5, and the probability that he will be in both the hurling and football teams is 0.3. His father says he will buy him a bicycle if he makes any of the school sport teams and he plays only those two sports at school. What is the chance that he gets a bicycle from his dad?

Note to HL learners: Revisit the questions above and try answering them with the formula.

Activity 3.3

Consider the following game

Players roll 2 four-segment spinners, which have equal segments numbered 1, 2, 3 and 4 . Player 1 wins if the sum of the spinner numbers is 3, 4, or 5,; player 2 wins if the sum is 2, 6, 7, or 8.

a. Predict whether player 1 or player 2 has the greater chance of winning. Play the game a few times to check your prediction. Now use the table below to help you decide in a more mathematical way. Write a sentence explaining why you think the game is, or is not, fair.

	1	2	3	4
1				
2				
3				
4				

b. Now consider this game

Players roll 3 four-segment spinners, which have equal segments numbered 1, 2, 3, and 4. Player 1 wins if the sum of the spinner numbers is 3, 4, 5, 6 or 12; Player 2 wins if the sum is 7, 8, 9, 10 or 11.

Is this game fair?

Can you represent the possible outcomes in the same way? It is difficult because there is an extra dimension – the 3rd spinner.

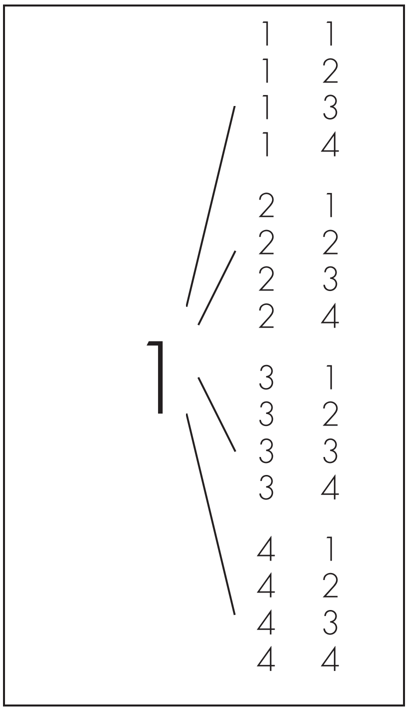
Consider all the possibilities when the first spinner shows a 1.

This is only ¼ the total number of outcomes and the process of completing the rest gets very repetitive.

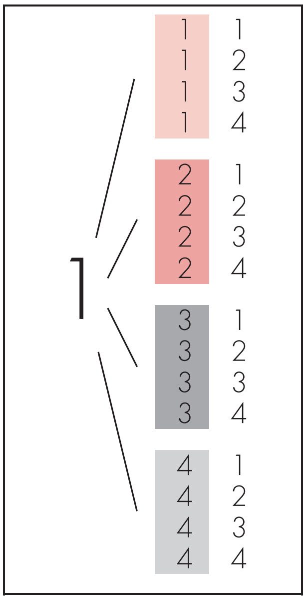
1	1	1
1	1	2
1	1	3
1	1	4
1	2	1
1	2	2
1	2	3
1	2	4
1	3	1
1	3	2
1	3	3
1	3	4
1	4	1
1	4	2
1	4	3
1	4	4

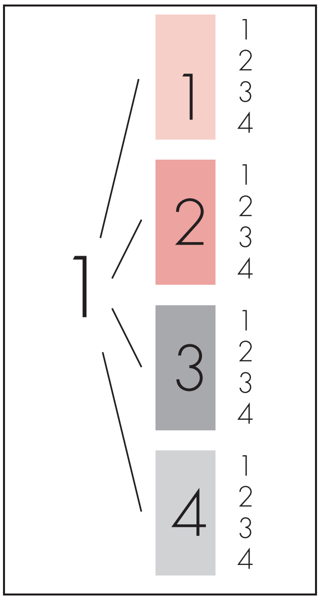
We could get rid of the repetitions by replacing the first column of 1's with 1 big 1.

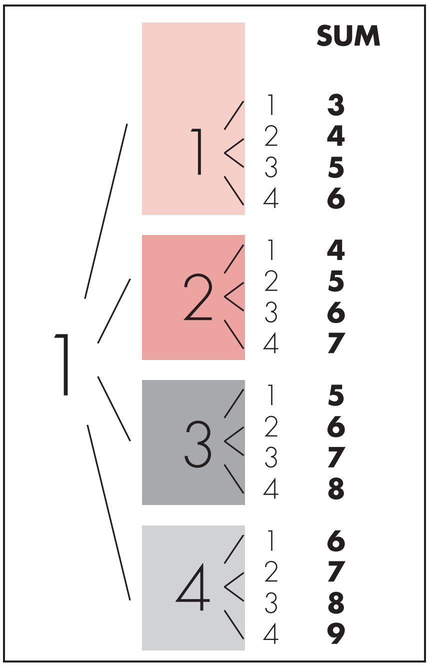
1	1	1
1	1	2
1	1	3
1	1	4
1	2	1
1	2	2
1	2	3
1	2	4
1	3	1
1	3	2
1	3	3
1	3	4
1	4	1
1	4	2
1	4	3
1	4	4



Can you get rid of any more repetitions?







Can you see a pattern forming ?
This is called a tree diagram; can you see why? Can you see how the required outcome (sum of the three spinners) is calculated for each 'branch' of the 'tree'?

i. Draw tree diagrams showing the possible outcomes when the first spinner shows 2, 3, and 4.
ii. How many possible outcomes are there? Now use your diagrams to decide if the game is fair (see the rules at the start).

This is how one student explained why tree diagrams are very useful when counting outcomes such as in this question:

> *Well, tree diagrams are useful for counting the total number of outcomes. There are four 'trunks' (for the possible numbers on the first spinner), and each has four 'branches' (for the possible numbers on the second spinner), and each has four 'twigs' (for the possible numbers on the third spinner). An outcome is formed as we go from a trunk to a branch to a twig. There are as many outcomes as there are twigs: 4x4x4=64.*

c. Draw a tree diagram showing the number of possible outcomes when three coins are tossed

Could you have answered this question without drawing the tree diagram?

Explain

i. Use your tree diagram to answer the following

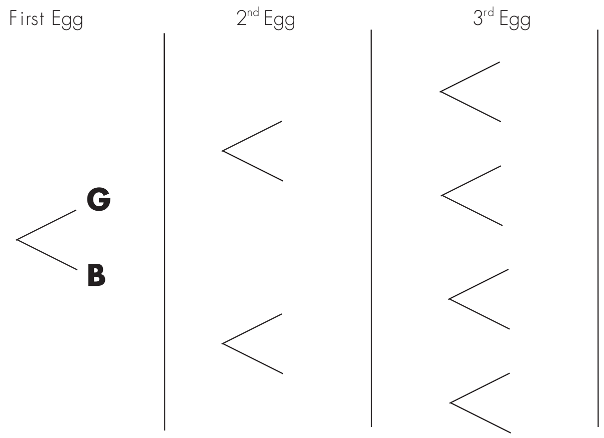
P (All tails) =

P (All heads) =

ii. Predict the number of possible out comes when two coins are tossed and 1 die is rolled. Check your prediction by drawing a tree diagram.

Q. 3 There are a dozen eggs in a box and 3 of them are 'bad'. 3 eggs are chosen at random from the box.

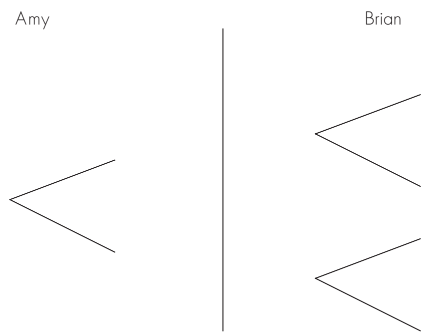
a. Complete the probability tree diagram below, showing good (G) and bad (B) eggs.



b. Work out the probability that
   i. all three eggs are 'good'
   ii. 1 egg is 'bad'
   iii. 2 eggs are 'bad'
   iv. all three eggs are 'bad'

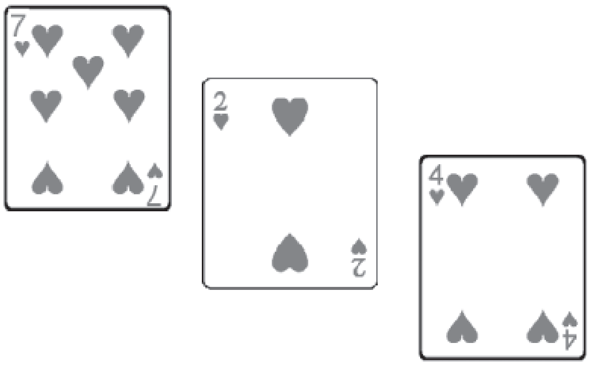
Q. 4 Jessica is taking part in a quiz. She is unsure of the answer to a question and needs to ask her team-mates, Amy and Brian. The probability that Amy will get it right is 0.7.The probability that Brian will get it right is 0.4.

a. Complete the probability tree diagram below.



   i. What is the probability that at least one of her two friends will give her the correct answer?
   ii. What is the probability that neither of them will give her the correct answer?

Q. 4 John and Sophie each have three cards numbered 2, 4 and 7. They each select one of their own cards. They then add together the numbers on the four remaining cards. What is the probability that their answer is an even number? Explain how you arrived at your answer.



Note to LC-HL learners: Revisit the question and answer it without the aid of a tree diagram.

Q. 5 Suppose that every child that is born has an equal chance of being born a boy or a girl.

i. Write out the sample space for the situation where a mother has two children.

ii. What is the probability that a randomly chosen mother of two children would have two girls?

iii. What is the probability that this mother of two children would have two boys?

iv. What is the probability that this mother of two children would have one boy and one girl?

Q. 6 Consider the situation where a die is rolled once; list the outcomes in the sample space.

The following events are defined:

A: a number greater than 2 but less than 5 is obtained.

B: an odd number is obtained.

i. Give the Venn diagram that depicts this situation.

ii. Find the following probabilities:

(a) P(A) (b) P(B) (c) $P(A \cap B)$ (d) $P(A \cup B)$ (e) $P(A^C)$

iii. Give a full account of what you understand by P(A) and $P(A \cup B)$.

Q. 7 A company that makes baby food tins finds the probability of producing a tin with a scratch on it to be 0.04, the probability of producing a tin with a mark on it is 0.07, and the probability of producing a tin with a scratch and a mark on it is 0.03.

a. Define events

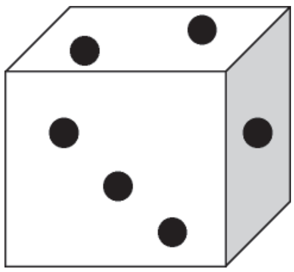
M: Tin produced with mark on it
S: Tin produced with scratch on it

i. Are these events mutually exclusive?

ii. Find the probability of producing a tin with a mark OR a scratch on it.

iii. How many tins, out of every 1000 produced, are expected to be scratched OR have a mark on them?

Q. 8 Judith is playing a game in which she moves a marker along a circular track. She uses a coin and a six-sided die to decide the steps she takes at each move.



She tosses the coin to see whether she moves forwards (if a head is shown) or backwards (if a tail is shown) and then rolls the die once to get the number of steps to be moved.

a. Complete a sample space to show all possible outcomes.

b. What is the probability that Judith will have to move

i. backwards fewer than 3 steps?

ii. forwards more than 4 steps?

Show how you arrived at your answer in each case by showing the steps involved.

Q. 9 A spinner with three equal segments numbered 1, 2 and 3 is spun twice and the number obtained each time is noted.

i. Draw a tree diagram to show all possible outcomes of the experiment.

ii. Write out the sample space of this experiment.

iii. Find the probability that the spinner stops on an odd number both times it is spun.

iv. Find the probability that the spinner stops on an odd number the first time it is spun and an even number the second time it is spun

Q. 10 For each of the pairs of events, A and B, determine whether the events are mutually exclusive and/or independent.

Event A	Event B	Mutually Exclusive	Independent
First born child is male	Second born child is female		
Coin displays head upon landing	Coin displays tail upon landing		
Die shows even number when tossed	Die shows number > 3 when tossed		
Card drawn is black	Card drawn is 3 of Clubs		
Ball chosen from bag containing different coloured balls	Second ball chosen from same bag without replacement of first ball		
First-born child has red hair	Second-born child has red hair		
Student will get A in mathematics exam	Student will get A in Irish exam		
Father is Bald	Eldest son will become bald		
Person abuses drugs	Person is left handed		
Person smokes cigarettes	Person dies of lung cancer		
Parent is a twin	Child is a twin		
Traffic Jam occurs on M50 between time T and T+1hour	Traffic jam occurs on the M50 between the time T+1hour and T+2hours		
Flooding occurs on July 2nd this year in Galway	Flooding occurs on July 2nd this year in Mayo		
Football team wins FAI cup	Football team wins League of Ireland Championship		
Player wins GAA hurling All Ireland medal	Player wins GAA camogie All Ireland medal		
Country possesses nuclear technology	Country is member of European Union		
A fair coin is tossed and 99 times in a row the coin displays heads	The 100th fair coin toss displays a head		

PROBABILITY 4

SYLLABUS TOPIC: COUNTING AND CONCEPTS OF PROBABILITY

LEARNING OUTCOMES

As a result of completing the activities in this section you will be able to

- list outcomes of an experiment
- apply the fundamental principle of counting
- count the arrangements of n distinct objects (n!)
- count the number of ways of selecting r objects from n distinct objects
- count the number of ways of arranging r objects from n distinct objects

INTRODUCTION

The activities are designed to help you build up an understanding of the concepts of arranging and combining numbers. Some of the activities will be done in class under the direction of your teacher; others can be done at home. Your answers to the questions will give you and your teacher an indication of how well you have understood the concepts.

Activity 4.1

Choose 7 classmates, friends or family members and ask them to stand in a row facing you. Write the numbers 1 to 7 on seven separate pieces of paper. One way to hand out the seven numbers to the seven people is to give the person on the left of the row the number 1 and the next person the number 2 and so on until the person on the right end of the row receives the number 7.

Alternatively you could hand out the numbers in decreasing order from left to right so that the person on the left receives the number 7 and the person on the right end receives the number 1.

This gives us two different **assignments** of numbers.

i. How many different assignments of the seven numbers are possible?

Make a table to help organise your work. Can you see any patterns in your work?

Place the seven pieces of paper in a bag and moving along the row of seven people ask each person in turn to choose one piece of paper without looking at the pieces of paper.

ii. What is the probability that the numbers were assigned in perfectly increasing order from 1 to 7 from the left of the row to the right of the row?

Ask everyone to show you the number that they got.

iii. Did the numbers line up perfectly in an increasing order from left to right?

Now choose the three tallest of the seven people and ask them to stand to one side in a new row.
Take the three pieces of paper from the three people.

iv. Are the numbers 1, 2 and 3 on the three pieces of paper?

v. Is it possible that the three pieces of paper could have the numbers 1,2 and 3?

vi. Is it possible they could have had the numbers 4,5 and 6?

vii. Is it possible that all three pieces of paper could have odd numbers?

viii. Is it possible all three could have even numbers?

ix. In total how many different sets of numbers could be on the pieces of paper that the group of three people have?

Ask each of the three people to choose one of the three pieces of paper that you took from them. (Make the three people close their eyes when they are choosing, so that their choices are random.)

x. Did the three people choose the same numbers that they had originally chosen?

xi. How many ways could the three people choose from among the three numbers?

Q. 1 Considering the experiment you have just performed, answer the following questions.

xii. How many ways can three people choose three numbers from the numbers 1-7 if all you are interested in is the set of three numbers that are chosen and you are not interested in who received which number? (LCOL)

xiii. How many ways can three people choose three numbers from seven numbers taking into account that you now do care about who receives which number? (LCHL)

Q. 2 Supposing now that instead of the numbers 1-7 we had the numbers 1-42,

i. how many ways could we choose three numbers from the forty two numbers assuming that the order of the three numbers is irrelevant?

ii. how many ways could we choose six numbers from the forty two numbers assuming the ordering of the numbers is irrelevant?

Activity 4.2

a. Get a pack of playing cards and remove the two jokers from the pack. How many cards are left in this pack? Choose a card at random from the pack and record on a piece of paper what the card is then replace the card in the pack and shuffle the cards. Repeat this until you have selected 100 cards in total.

i. Compute the relative frequency of cards that were chosen which were red cards.

ii. Compute the relative frequency of cards that were chosen which were hearts.

iii. Compute the relative frequency of cards that were chosen which were both red and hearts.

You may find it helpful to create a table in which to record your results. Look back at previous tables you have used.

b. You are now going to calculate some theoretical probabilities. Each time we calculate a probability we must first determine the sample space.

Consider choosing 1 card from the above pack of cards.
What is the sample space for this experiment? How many elements are in the sample space?
Now consider some different events that we could choose from this sample space.

i. How many ways can you choose a heart from the pack of cards?

ii. How many ways can you choose a red card from the pack of cards?

Dividing the number of elements in each event by the number of elements in the sample space compute, the probability of

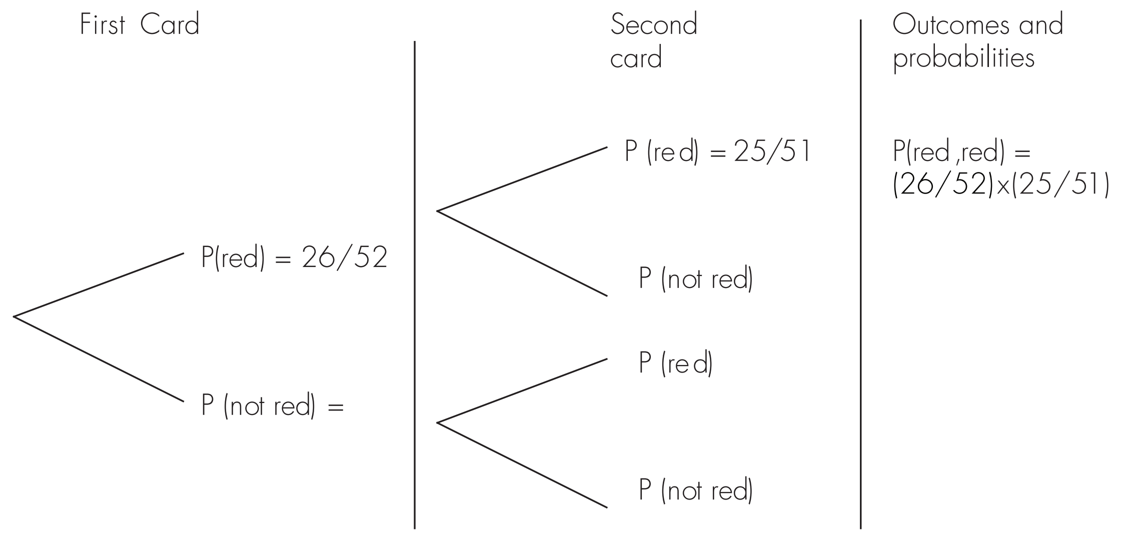
iii. choosing a heart from the pack of cards

iv. choosing a red card from the pack of cards

Activity 4.2A

So far you have chosen only 1 element from the sample space. Now you will progress to more challenging examples where you are required to choose more than 1 element from the sample space

Complete the tree diagram and use it to find the probability of choosing two red cards from a pack of 52 cards. Now realising that cards that are not red are black, what is the probability of choosing two black cards?



What is the probability of choosing one red card followed by one black card?
What is the probability of choosing one black card followed by one red card?
Hence, by adding probabilities, calculate the probability of choosing one red and one black card from a pack of 52 cards.

Activity 4.2B

You can approach this question a different way.
Consider choosing 2 cards from the above pack of cards. Think about the sample space for this experiment. How many elements are in the sample space? (In other words, how many ways can you choose 2 objects from a set of 52 objects? In mathematical language this is written as ${}^{52}C_2$.)
Now consider some different events that we could choose from this sample space.

> If you are having difficulty answering this question try to simplify it by using smaller amounts of cards e.g. how many ways can you choose 2 cards from a pack of 3 different cards (${}^{3}C_2$) or how many ways can you choose two letters from A, B, C (${}^{3}C_2$)?. In this way you may see some patterns in your work that will help when you use larger numbers.

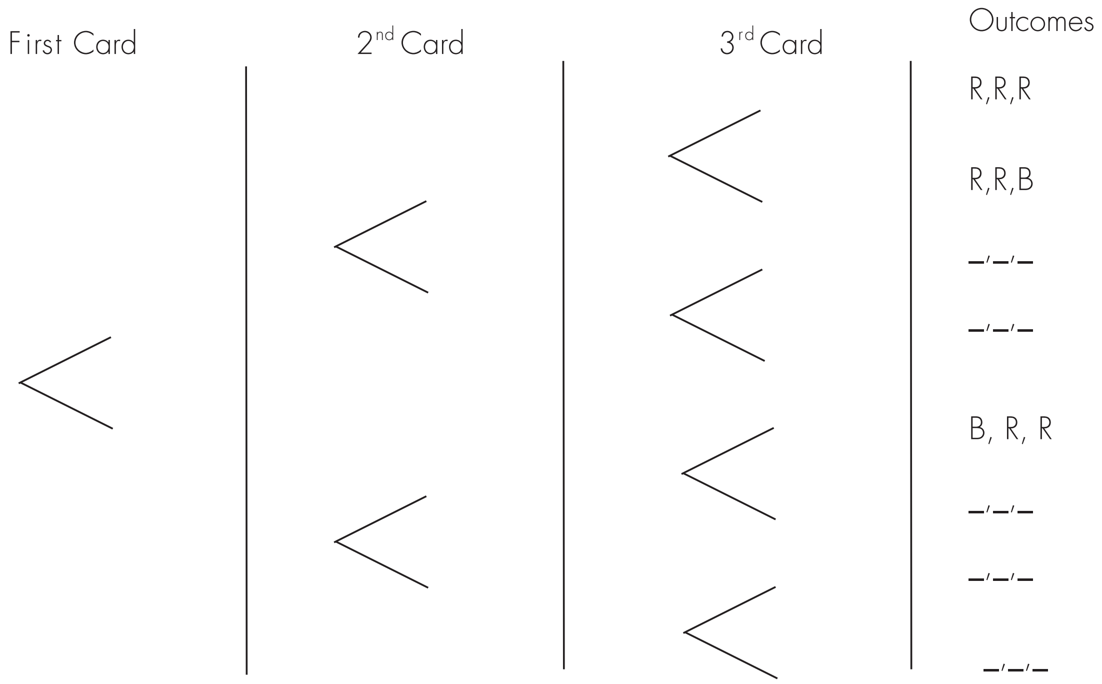
i. How many ways can you choose two red cards from the pack of cards? (When answering this question first consider how many red cards there are and then consider how many ways to choose two of these)
ii. Dividing the number of elements in this event by the number of elements in the sample space compute the probability of choosing two red cards from the pack of cards

> You have now answered the same question in two different ways; compare this answer to the answer that you obtained earlier.
>
> Show that the answers you got from these two methods are actually the same answer (you will need to review the definition of combinations and remember how to divide fractions).
>
> Which approach to answering this question did you find the most helpful?

Now, by counting the number of elements in the events and in the sample space as above what is the probability of choosing two black cards? Compare this answer with the one you got using the tree-diagram approach.

Activity 4.2C

We have seen questions involving choosing one and then two cards from a pack of cards. Now imagine choosing three cards from a pack of 52 cards. Using the tree diagram approach find the probability of choosing three red cards from a pack of 52 cards.



Now consider answering this question by counting the number of elements in the sample space and counting the number of elements in the event (i.e. that there are three red cards). Firstly, how many ways can you choose three cards from a pack of 52 cards? This is the size of your sample space. Now how many ways can you choose three red cards from 26 red cards? This is the number of elements in your event. Now divide the number of elements in the event by the number of elements in the sample space; what is your answer? This is the probability of choosing three red cards from the pack of 52 cards.

Show that the answers you got using the tree diagram method and the sample space method are identical.

Now consider choosing two red cards and one black card from the pack. How many ways can you do this?

Look back at your work from the previous section.
What is the probability of getting a black card first followed by two red cards?
What is the probability of getting a red card first followed by a black and then another red?
What is the probability of getting two red cards followed by a black card?
Think about the sample space for this experiment. How many elements are in the sample space? (In other words, how many ways can you choose 3 objects from a set of 52 objects? In mathematical language this is written as ${}^{52}C_3$.)
Now consider some different events that we could choose from this sample space.

Compare the answers you obtain from both approaches.

Q. 1 Consider choosing two cards from a pack of 52 playing cards.

i. How many ways can you choose two cards from the pack of cards?
ii. How many ways can you choose two Queens from the pack of cards?
iii. What is the probability of getting two Queens?
iv. What is the probability of getting one Queen and one King?
v. How many ways can you choose five cards from a pack of cards?
vi. What is the probability that you get four Kings and 1 queen in a hand of five cards?
vii. What is the probability that you get three Kings and Two Queens in a hand of five cards?

Construct a mind map or graphic organiser to help you remember how you

- count the arrangements of n distinct objects (n!)
- count the number of ways of selecting r objects from n distinct objects
- count the number of ways of arranging r objects from n distinct objects

Activity 4.3

Find out how the National Lottery game of Lotto is played.

a. How many different sets of six balls could possibly be chosen in the Lotto game?

> Can you see any similarity between how you approach this question and the work you have done previously. In activity 4.2 you saw two approaches to answering these questions How many branches would you have in your tree diagram? Is it wise to attempt to write out all the combinations? Is the tree diagram going to be too complicated for this example?

b. Suppose that you wrote down six numbers from among the set of numbers 1 to 45. What is the size of the sample space? (How many ways can you choose 6 numbers from 45 different numbers?)

c. What is the probability that the numbers you wrote down would be the exact same numbers chosen in next Saturday's Lotto game?

d. How many ways could you choose a set of five numbers from the six numbers that are chosen in the Lotto game next Saturday?

> This is complicated. Think how it relates to the card situation in activity 4.2. Consider that the red cards are the winning 6 numbers chosen in the draw. The black cards are the numbers not chosen in the draw. This gives you a pack of 45 cards with 6 of them red and 39 of them black.

e. How many numbers will NOT be chosen in the Lotto game on Saturday?

f. How many ways could you choose one number from among this set of numbers which will NOT be chosen in the Lotto on Saturday?

g. Combining the answers to the last three questions, can you establish how many ways EXACTLY five of the six numbers that you wrote down match with five of the numbers that will be chosen on Saturday's Lotto draw?

Suppose the Lotto game was structured so that you chose 6 numbers from the numbers 1-36. Would that be an easier game to win?

Can you calculate the probability of matching all 6 numbers in a Lotto game that contains 36 numbers? What about one that contains 39 numbers and one that contains 42 numbers?

Would it be good for people playing to decrease the number of numbers in the Lotto game from 45 to 36? Can you predict what would happen to the number of winners if this were to happen?

PROBABILITY 5

SYLLABUS TOPIC: CONCEPTS OF PROBABILITY

LEARNING OUTCOMES

As a result of completing the activities in this section you will be able to

- calculate expected value and understand that this does not need to be one of the outcomes
- recognise the role of expected value in decision making with a focus on fair games
- change the rules of a game to make it 'fair'
- invent games for your peers
- engage in discussions about why car insurance premiums are more costly for some

INTRODUCTION

The activities described below and the questions that follow allow you to construct an understanding of expected value and how the expected value can be used to inform decision making. You will recall your work on deciding whether a game is fair or not and extend this to understand how calculating the expected value facilitates these types of decisions. The activities are grounded in context so that you will not only appreciate how the mathematical concepts relate to real life situations; but also appreciate the usefulness of mathematics in gaming, the insurance and financial industry.

Prior knowledge

You will need to be able to use sample spaces and tree diagrams to calculate the probability of events occurring.

Note: Sometimes it is desirable to know the mean or average outcome of an experiment; in the long run this is called the **expected value**.

Read and reflect on the following definition

In probability theory the expected value of a discrete random variable is the sum of the probability of each possible outcome of the experiment multiplied by the outcome value. Thus, it represents the average amount one 'expects' as the outcome of the random trial when identical odds are repeated many times. Note that the value itself may not be expected in the general sense - the 'expected value' itself may be unlikely, or even impossible.

This has been summarised below

Expected Value calculations essentially involve taking all the possible outcomes, weighting the more likely outcomes, and coming to a conclusion. It is an average of all the likely outcomes.

We will consider some examples from real life where it is desirable to know the expected value. The most easily understood examples are from gaming. Consider the following

Ann and Barry each have a pile of sweets and they play a game:
Barry rolls a die. If he gets a 6, Ann gives him four sweets; if he gets a 1, he gives Ann two sweets; if he gets a 2, 3 or 4, he gives Ann one sweet; if he gets a 5, there's no exchange of sweets.

Is this a fair game, or does it favour one player over the other?
If it's not fair, how can it be changed to make it fair?
Make up some similar games that are fair /unfair

Think:

What constitutes a 'win' for Barry?

What constitutes a 'loss' for Barry?

Complete the **Before** section of the prediction sheet below

Before		Statement	After	
Agree	Disagree		Agree	Disagree
		1. Playing this game requires skill.		
		2. Barry will usually win when he plays this game.		
		3. Every time Barry plays this game, he will have the same chance of losing.		
		4. This is a fair game.		

What is the probability that Barry will win?

What is the probability that Barry will lose?

What is the 'value' for each outcome? For example, if he throws a 6 he will get four sweets (i.e. he is 'up' 4) so the value would be +4.

Complete the following table by listing all the possible outcomes, calculating the corresponding probabilities and inserting the 'value' for each outcome

Possible Outcome	Probability	Value

Look back at the definition try to work out how to calculate the expected value from the table.

......the expected value is the sum of the product of probability times the value....

So what is the expected value in this game?
What does this mean?
Is the game fair?
Who does it favour?
If the expected value was positive who would it favour?
If the game is fair what should the expected value be?

Refer back to the table; complete the **After** section. This will give you an indication of how well you have grasped the underlying concepts.

Would the game be fair if, when Barry throws a 5, Ann gives him one sweet?
Justify your reasoning

Reflect on how you calculated the expected value.

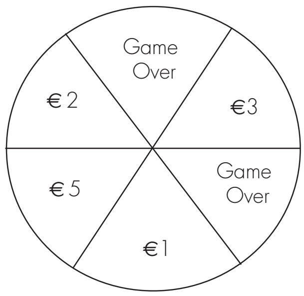
Create a mind map or a graphic organiser that will help you remember how to calculate the expected value (http://www.action.ncca.ie).

Q. 1 Spin and Win

Instructions

- one game costs €2 to play
- spin and win the amount shown
- Tick Agree or Disagree in the Before column beside each statement before you start to play the Spin and Win game.
- Revisit your choices at the end of the investigation.

Before		Statement	After	
Agree	Disagree		Agree	Disagree
		1. Playing this game requires skill.		
		2. You will usually win when you play this game.		
		3. Every time you play this game, you will have the same chance of losing.		
		4. This is a fair game.		

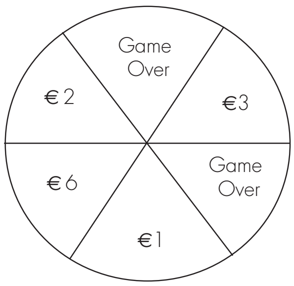


Think: How can what you have learned in the previous example help you make a decision about whether or not this game is fair?

Look back at how you assigned values in the previous example.
What is the difference in this game? Did Barry or Ann have to pay any sweets to play the game?
Do you have to pay to play *Spin and Win*? Will this affect the 'value' you assign to a particular outcome?
Is this a fair game? Justify your answer

Now return and complete the **After** column of your table.

b. Consider the following statements. What effect, if any, will they have on your decision about whether or not the game is fair? Justify your answer.
   i. The cost of the game is increased to €3
   ii. The following spinner was used



Return again to your graphic organiser and adjust it to take into consideration what you have learned from playing ***Spin and Win***

Q. 2 You and a friend are playing the following game:

Two dice are rolled. If the total showing is a prime number, you pay your friend €6. Otherwise, your friend pays you €2.

i. What is the expected value of the game to you?

Remember: ***expected value is a weighted average of the values of the outcomes. List all the outcomes; multiply the value of each outcome by the probability of that outcome, then add.***

ii. If you played the game 40 times, what are your expected winnings?
iii. After playing the game a while, you begin to think the rules aren't fair and you decide to change the game. How much (instead of €6) should you pay your friend when you lose so that your expected winnings are exactly €0?

Q. 3 A roulette wheel has 38 numbers; 18 black, 18 red and 2 green

a. Game1: *Bet on a Number*

Bet €1 on a number; if you are right you win €35 (and you get your €1 back). The outcome is +35 because you are 'up' €35. If you are wrong, you lose €1. What is your outcome?
Is this a fair game?

b. Game 2: *Bet On Black*

Bet €1 on a black number
If right win €1 (and get your €1 stake back). If wrong, you lose €1.
What is the expected value? Is this a fair game?

The following example is more challenging, since it doesn't refer to gaming.

Q. 4 Consider two family planning strategies.

(a) have children until you have a girl or (b) stop after two boys.
Would either strategy upset the male /female mix in the population? What do you think? Make a prediction.

If you calculate the **expected value** or mean number of boys for a large population of families and compare this with the **expected value** or mean number of girls for a large population of families you will be able to answer this question.
Discuss in your group why this information will help you decide if either strategy would upset the male/female mix.

Use probability trees to help find the probability of each outcome. Think about the possible outcomes
Girl first (what happens now?); Boy, Girl (what happens now?); Boy, Boy (what happens now?)
What's the average number of boys per family? Complete the probability column in the table below.

Now look at the definition of expected value again and see if you can calculate the expected value or the average number of boys per family. Can you complete the final column in the table?

Boys	Probability	Value × Probability
0		
1		
2		

So, the Expected Value, or average number of boys per family =
Hence, the mean number of boys per family for a large population of families =

Now in a similar way calculate the average number of girls per family.
Is there any difference in these values?

Decide whether either strategy will affect the mix of boys and girls in the population?
Was your prediction correct?

Thoughts
Think about how bookmakers and casinos ensure that they make a profit – by making sure that the expected return for the player is always lower than the stake. Players will of course occasionally win, but the house wins in the long run (average over all players and all games).

Insurance is mathematically identical to gambling against a casino. For example, when you pay your life insurance premium each month, you're betting that you're going to die this month and the insurance company is betting that you won't! In game-theory terms, insurance is a bet that does not favour the player, which leads to the question about why you would have it, and this could lead on to the idea of 'risk'. There's more to decision-making than just looking at the expected value. The bottom line here is that when sums of money or other consequences are very large, their true value is not really in proportion to their actual face value. For example, one might judge the need for their family to get half a million euro if they die this year to be more than a thousand times more valuable to their lives than the €500 loss they suffer by buying the insurance. The insurance company deals with sums of money on this scale as being in proportion to their face value, but to the individual it's the impact on your life that counts. (This kind of issue can also provide interesting discussion on how the optimal strategy for 'Who wants to be a millionaire?' may be different for people in different financial circumstances.)

Think about expected value and become an informed citizen.

Why are car insurance premiums higher for some categories of driver?
What are the advantages/disadvantages of community rating in the health insurance market?
Why is a standard retirement age of 65 in an era of increasing longevity a pensions/actuarial nightmare?

PROBABILITY 6

SYLLABUS TOPIC: CONCEPTS OF PROBABILITY; CONDITIONAL PROBABILITY

LEARNING OUTCOMES

As a result of completing the activities in this section you will be able to

- solve problems involving conditional probability in a systematic way
- extend your understanding of the basic rules of probability (AND/ OR, mutually exclusive) through the use of formulae, in particular
  $P(A \cap B) = P(A) \times P(B|A)$

You will also examine specific situations that will help you to

- appreciate contexts where $P(A | B) \neq P(B | A)$

INTRODUCTION

The activities described below and the questions that follow give you the opportunity to reinforce your understanding of the basic concepts of probability and to construct an understanding of the concept of conditional probability.

Activity 6.1

Consider a dice game with two six-sided dice. The sum of the dice decides who wins. What is the probability that the sum of the numbers on the dice is seven?
Use the table below to help make your decision.

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

How many possible outcomes are there when you roll the 2 dice?

Consider two events

Event A = {The sum of the numbers the dice is 7 or 9}.
Event B = {The second die shows 2 or 3}

Compute P(A) and P(B) (the probabilities of events A and B)

Use the table below to help work out P(A)
Colour the table to show all the possible outcomes for event A
Now find P(B). Use a different colour and show all the possibilities of event B

Are there any outcomes that are common to both?
If so, how many?

Now use your shaded table to find P(A|B) ie the probability of the sum of the numbers the dice shows is 7 or 9 and the 2nd die shows 2 or3

Use this result to write a general equation for P(A|B) in terms of

i. possible outcomes,and
ii. probabilities

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

Activity 6.2

For this experiment you will need ten girls and ten boys and three rooms. If you do not have access to ten boys and ten girls write the word 'boy' on ten pieces of paper, the word 'girl' on ten other pieces of paper; then use three paper bags to represent the three rooms.

a. Place all ten boys and ten girls in Room 1. Now choose four girls and six boys and put them in a Room 2. Leave the room and place a blindfold over your eyes and re-enter Room 2. Walk around until you encounter a person. Ask that person to leave Room 2 and go to Room 3. Continue walking around while wearing your blindfold until you bump into a second person. Ask this person to leave the room and go to Room 3.

Go to Room 3 and determine whether there are two boys, two girls or one boy and one girl in Room3.

Now start again from the beginning and repeat this experiment 20 times. Each time record the gender of the two people who are in Room 3.

Compute the relative frequency for each of the events:

- A Room 3 contains two girls
- B Room 3 contains two boys
- C Room 3 contains a boy who was chosen first and then a girl who was the second person chosen
- D Room 3 contains a girl who was chosen first and also a boy who was the second person chosen

Now let's calculate the theoretical probabilities of each of these events. Recall that in Room 2 there are four girls and six boys.

- i. What is the probability that the first person chosen from Room 2 is a boy?
- ii. Suppose that the first person chosen was a boy, how many boys and how many girls are left in Room 2?
- iii. Now what is the probability that the second person chosen from this remaining set of girls and boys in Room 2 is a boy?
- iv. Combining these two probabilities appropriately what is the probability that the two people chosen to enter Room 3 were both boys?
- v. Using the same logic, compute the probabilities of events A, C and D. Are the probabilities of C and D the same?
- vi. What is the probability that Room 3 at the end of the experiment contains one boy and one girl? (You do not know the order in which the people were chosen to enter Room 3.)

b. Put all of the twenty people back in Room 1 and this time place five boys and five girls in Room 2. Now wearing your blindfold enter Room 2 and choose one person at random. Ask this person to enter Room 3 and write on a blackboard in Room 3 whether they are male or female and then return to Room 2. Now repeat this process again.

Compute the probability of the following events:

A. The first person to have been chosen was a girl.
B. The first person to have been chosen was a boy.
C. The second person to have been chosen was a girl.
D. The second person to have been chosen was a boy.
E. The first person chosen was a boy and the second person chosen was a girl.
F. The first person chosen was a girl and the second person chosen was a boy.
G. The first person chosen was a boy and the second person chosen was a boy.
H. The first person chosen was a girl and the second person chosen was a girl.

i. Examining these probabilities determine whether the events A and C are independent. What about A and D?

ii. How do the probabilities for events E, F, G and H compare? Does it matter in which order the boys and girls were chosen?

c. Suppose the event X was that a boy is chosen from Room 2 and event Y is that a girl is chosen.

iii. What is $P(X|Y)$ – the probability that a boy is chosen given that a girl was already chosen?

iv. What is $P(Y|X)$ – the probability that a girl is chosen given that a boy was already chosen?

d. Suppose a third experiment was performed in which we again start with 5 boys and 5 girls in Room 2 but this time when someone is chosen to enter Room 3 they stay in Room 3 as they did in our first experiment.

i. Under these rules, what are the probabilities of the events A, B, C, D, E, F, G and H?

ii. Are A and C independent in this experiment?

iii. What about A and D?

iv. How do the probabilities for events E, F, G and H compare now?

v. What is $P(X|Y)$ – the probability that a boy is chosen given that a girl was already chosen?

vi. What is $P(Y|X)$ – the probability that a girl is chosen given that a boy was already chosen?

Let Z be the number of boys who could be in Room 2 at the end of the last experiment.

Think: What values can Z possibly take?

Looking at the probabilities you have previously calculated, assign a probability to each possible value of Z. Now compute E(Z), the expected number of boys in Room 3. Compute E(Z) for the two experiments described earlier in this example.

Q. 1 You may remember doing the first part of this question before. Now you will extend the concept to consider conditional probability.

Suppose every child that is born has an equal chance of being born a boy or a girl.

i. Write out the sample space for the experiment where a mother has two children.
ii. What is the probability that a randomly chosen mother of two children would have two girls?
iii. What is the probability that this mother of two children would have two boys?
iv. What is the probability that this mother of two children would have one boy and one girl?

Suppose you meet someone who tells you they have two children and that one of their children is a girl.

v. What is the probability that this person has two daughters?
vi. Is it the same as it was in the previous situation of the randomly chosen mother?

Consider the sample space that you wrote out in the case of the randomly chosen mother.

vii. Do all of the elements of this sample space still apply to the example we are now considering?
viii. Write out the sample space for this example, i.e. where the person tells you that one of their children is a girl.
ix. Now, considering this sample space again, try to answer the question 'What is the probability that this person has two daughters?'

Extension
What if this person tells you that they have two children and that one of their children is a *girl named Florida*? Does this extra piece of information make any difference? If you have access to the internet, you might look up this well known puzzle/problem, which is not as straightforward as it might first appear.

Q. 2 100 glasses of mineral are placed on a table. One of these glasses of mineral contains a new brand of cola called Green Cow Cola and the other 99 are a well known brand of cola.

a. A student is asked to go into the room and choose one of the glasses at random not knowing in advance which glass contains the new brand of cola. The student considers themselves a connoisseur of cola and is able to correctly identify colas 95% of the time. This means that 95% of the time if the student is given any brand of cola in a glass and is asked to identify it they will assign the correct name to that cola.
   i. What is the probability that they will choose the glass of Green Cow Cola?

ii. What is the probability they will not choose the glass of Green Cow Cola?

Let A be the event that a glass contains Green Cow Cola.

iii. What, in ordinary English, is the event A^C, the complement to A?

Let B be the event that the student says that a glass contains Green Cow Cola after they have tasted the cola.

iv. What is the event B^C?

v. What is the $P(B|A)$ – the probability that the student says the glass contains Green Cow Cola given that the glass actually contains Green Cow Cola?

vi. What is $P(B|A^C)$?

vii. What is $P(B^C|A)$?

viii. What is $P(B^C|A^C)$?

ix. In ordinary English describe what $P(A|B)$. Do you think it is the same as $P(B|A)$?

x. What is $P(A \cap B)$ in English? What is $P(B \cap A)$ in English? How do these two probabilities compare?

b. Write down formulae to express each of $P(A|B)$ and $P(B|A)$ in terms of $P(A)$, $B(B)$ and $P(A \cap B)$.

Considering these formulae and making an appropriate substitution from one formula into the other can you see how to express $P(A|B)$ in terms of $P(B|A)$?
Looking at the new expression you have created, once again try to answer the question: 'Is $P(A|B)$ the same as $P(B|A)$?'
Utilising the expression you have created can you compute a value for $P(A|B)$?

If you are having difficulty answering this question it may help to consider smaller numbers; in this way you will organise your work in a systematic way and you may see patterns that will help you answer questions with larger numbers. It may also help to draw a diagram to represent the situation

Activity 6.3

A disease called Stats-itis is known to affect one in every 100,000 people in the world. People who have this disease have a natural proficiency for Statistics and many of them eventually become employed as Statisticians.
There is a test that has been developed which can test teenagers to determine if they suffer from Stats-itis. While this test is not completely accurate it does have an accuracy that is comparable to tests used for many other well known diseases. If the test is performed on someone who suffers from Stats-itis there is a 0.95 probability that the test will correctly

indicate that the person has the disease. If the test is performed on someone who does not have the disease then the test will correctly identify that the person does not have the disease 95% of the time.

Suppose that you get tested for this disease and the test says that you have the disease. Recalling the 95% accuracy of the test, does your intuition tell you that it is pretty certain that you do indeed have the disease?

Let's now try and consider this question from an in-depth mathematical point of view and see how the true answers we receive compare with our intuition.

a. Let A be the event that the test indicates that someone has Stats-itis. Let B be the event that someone actually has Stats-itis.

   i. What is $P(A|B)$?
   ii. What is $P(B)$?
   iii. Using these two answers compute a value for $P(A \cap B)$ and for $P(B \cap A)$.
   iv. Draw a Venn-diagram for this scenario and identify the sets A, A^C, B and B^C.
   v. The following equation is missing one part. Can you fill in the missing part?

      $P(A) = P(A \cap B)$ + [missing part]

   vi. Can you express $P(B|A)$ in terms of the probabilities of the events A, B and 'the intersection of A and B'?
   vii. Express $P(A \cap B)$ in terms of $P(A|B)$.
   viii. Can you find similar expressions for $P(A \cap B^C)$, $P(B \cap A^C)$ and $P(A^C \cap B^C)$?
   ix. By combining several of the results you have already computed, express $P(B|A)$ in terms of $P(A|B)$.

b. Suppose you do not know whether you have Stats-itis.

   i. What is the probability that the test says you have Stats-itis given that you do have Stats-itis?
   ii. What is the probability that you have Stats-itis given that the test says you do have Stats-itis?

Suppose that you went to a doctor and the doctor performed this test on you and told you that you had the disease. Should you be concerned?

c. You ask the doctor how sure she is that you have the disease and the doctor says that the test is 95% accurate. The doctor seems to be indicating that she is 95% sure that you have the disease.

   i. Is this an accurate reflection of the true probability that you have the disease given that the test has come back positive?

   ii. How does the probability that you have the disease before you took the test compare with the probability that you have the disease given that you have taken a test and it has given a positive result indicating that you have the disease?

Consider this example in detail and discuss it with your friends and/or family. Does the answer you received for the 'probability that you actually have the disease given that the test says you have the disease' agree with your intuition and that of your friends or family?

Q. 2 Consider the following pairs of events:

1. A: A creature has a Tail.
   B: A creature is an Elephant.

2. A: A creature has wings.
   B: A creature can fly.

3. A: You are someone who plays a game of chance.
   B: You are someone who wins a game of chance.

For each pair of events answer the following questions:

i. What in words is $P(A|B)$?

ii. What in words is $P(B|A)$?

iii. Is $P(A|B) = P(B|A)$?

Q. 3 Consider the following events:

A: A person wears a black hoodie and blue jeans.
B: A person commits a crime.

John was arrested for stealing a woman's handbag. At the trial it is pointed out that a Garda arrested John 20 minutes after the crime took place 100 metres from the scene of the crime and that John was wearing blue jeans and a black hoodie.

i. Is $P(A|B) = P(B|A)$?
ii. How is this relevant to this court case?
iii. Is $P(A)$ a useful piece of information in establishing the innocence or guilt of John?
iv. Would it make any difference if this crime took place in the centre of Dublin compared to if the crime took place in a village with a population of 50 people?
v. Can you see any similarity between this question and the previous question about Elephants?

Evidence is then presented that the DNA matches an individual named Tom. The prosecution does not feel the need to present any other evidence relating Tom to the crime as they have the DNA match.

vi. Does your intuition tell you that this evidence is sufficient to establish with certainty that Tom committed the crime?
vii. By considering the total population of the world to be 6.75 billion, what is the number of individuals whose DNA would match the DNA from the crime scene?
viii. What is the probability that someone other than Tom could be the individual whose DNA was left at the crime scene?

PROBABILITY 7

SYLLABUS TOPIC: OUTCOMES OF RANDOM PROCESSES

LEARNING OUTCOMES

As a result of completing the activities in this section you will be able to

- apply the principle that, in the case of equally likely outcomes, the probability is given by the number of outcomes of interest divided by the total number of outcomes (examples using coins, dice, spinners, urns with coloured objects , playing cards, etc.)
- solve problems involving experiments whose outcome is random and can have two possibilities (labelled 'success' or 'failure'), such as tossing $n \leq 3$ coins or rolling $n \leq 3$ dice
- toss n coins or roll n dice and count number of 'successes' and calculate the probability of this occurring
- toss a coin or roll a die until the kth 'success' and calculate the probability of this occurring

Activity 7.1

Consider tossing three coins. Each coin has one face with a Head and one with a Tail. Coins never land on their side but always land face up.

Drawing a tree diagram will help you organise your thoughts and answer the following questions.

a. What is the probability that the first coin shows a Head?

b. What is the probability that all three coins show Heads?

Consider the event that two of the three coins were Heads and one was a Tail. Is it certain that the first two coins were Heads?
Is it possible that the last two coins were Heads?

c. What other possibilities are there?

If you are having difficulty answering this part of the question, write the numbers 1, 2 and 3 on three pieces of paper and place them in a bag. Draw three empty boxes in a horizontal row on a piece of paper. Number the boxes 1, 2 and 3 from left to right. Choose two pieces of paper from the bag and write the word Head in the boxes with the numbers on the two pieces of paper that you drew from the bag. Write the word Tail in the remaining box.

Place the pieces of paper back in the bag and repeat this experiment many times.
How many DIFFERENT possibilities are there for the arrangements of the two Heads and one Tail?
How many ways can you choose two numbers from the set of numbers 1, 2 and 3?

Activity 7.2

Imagine now that you are tossing four coins and consider the event that two of the coins show up as Heads and two as Tails.

i. Can you list all the possible arrangements of two Heads and two Tails among the four coins?

ii. How many different possibilities are there?

iii. How many ways can you select two objects from among four objects?

Q. 1 Suppose you were to toss one hundred coins and twenty three were to land as heads. It could be that the first twenty three coins were heads and the last seventy seven were tails. There could be other possibilities as well. How many ways can you select twenty three numbers from among the set of numbers 1 to 100? So how many different possible ways could you get twenty three heads among the one hundred coins?

Activity 7.3

a. You perform an experiment in which you will toss six coins, each time getting a Head or a Tail.

i. What is the probability that an individual coin toss results in a Head?

ii. What is the probability that an individual coin toss results in a Tail?

You may find it useful to draw tree diagrams to work out the probability for the next few questions.

iii. Is the event that you get a Head on the first coin independent of the event that you get a Tail on the second toss?

iv. What is the probability that the first two tosses reveal a Head on the first coin followed by a Tail on the second coin?

v. What is the probability that the first three tosses give exactly the pattern (Head, Tail, Tail)?

vi. What is the probability that the first coin shows a Head and all of the subsequent five coins show Tails?

vii. How many ways can you select one object from 6 objects?

viii. How many other arrangements of the coins would give one Head and five Tails?

ix. What is the probability that the coins land with the first five coins showing Tails and the last coin showing a Head?

x. What is the probability that the coins land with the second coin showing a Head and all the other coins showing Tails?

xi. By adding up the individual probabilities of each possible way that you can get one Head and five Tails, compute the probability that you would get exactly one Head when you toss six coins.

b. Now imagine repeating this experiment where you toss six coins, but suppose this time you get two Heads and four Tails. One way this could happen is that the first and third coins were Heads and the other four were Tails. What is the probability of getting exactly this arrangement? In total how many different ways can six coins display two Heads and four Tails? Are all of these different arrangements equally likely or is any one more likely than the others? What is the probability of each of these arrangements? Combining the information you have for the number of ways that you can get two Heads and four Tails with the probability of each of the possible arrangements, write down an expression for the probability of getting exactly 2 Heads when six coins are tossed.

   a. How many ways could you get six Heads among the six coins?

   b. How many ways could you get three Heads among the six coins?

Activity 7.4

a. Get a 2 Euro coin with a Harp on one side. Toss this coin and observe whether it lands with the Harp facing upwards. If it does, stop tossing the coin. If it does not, and instead you get a map of Europe displayed then toss the coin again repeatedly until the coin first displays a Harp. Count the number of tosses that were required until the first Harp appeared. Repeat this experiment 100 times (each class group could complete the experiment a number of times and compile the class results) each time noting the number of tosses until the first Harp appeared.

Construct a table in which to record your results

b. Draw a histogram of the relative frequencies of N where N is the 'Number of tosses required to get the first Harp'.

   i. Is there any pattern to the histogram?

   ii. What is the ratio of the relative frequency for N=3 versus the relative frequency for N=2?

c. You will now attempt to construct a probability distribution for this experiment.

What is the largest possible value that N can take? What is the smallest value that N can take?

If N=1 what happened when you tossed the coin?
If N=3 what happened each time you tossed the coin?

iii. What is the probability that you get a 'Europe' on the first toss of the coin?
iv. What is the probability that you get a Harp on the second toss of the coin?
v. Are the events 'Europe on first toss' and 'Harp on second toss' mutually exclusive and/or independent?
vi. What is the probability that you get a Europe on the first toss of the coin AND you get a Harp on the second toss of the coin?
vii. What is the probability that N=2?
viii. What is the probability that N=4?

If N=17, how many Europes have occurred?
If N=17, how many Harps have occurred?

ix. What is the probability that you get 25 Europes displayed one after another?
x. What is the probability that N=26?

Activity 7.5

Roll a standard six-faced cuboid die. (A die which has six faces displaying differing numbers of dots on each side from one dot to six dots.)

i. How many different possible numbers of dots can be displayed on the topmost side when the die lands?
ii. What is the sample space for this experiment?
iii. What is the probability that you get a 1 on the first roll? (The die falls with one dot facing upwards.)
iv. If you got a 6 on the first roll does that make it more or less likely that you will get another 6 on the second roll?
v. What is the probability that you will get a 6 on the second roll given that you have already gotten a 6 on the first roll?
vi. What is the probability that you get a 6 on the first roll and get a 6 on the second roll?
vii. What is the probability that you will not get a 6 on the first roll?

Now we will play a game in which you must roll the die as many times as is necessary until you get a 6. Record, N, the number of rolls required until you get a 6.

viii. What is the probability that you will not get a 6 on the first two rolls and you will get a 6 on the third roll?

ix. If N=67 what do you know about the 27th roll?

x. If N=34 what do you know about the 34th roll?

The table below describes the probability distribution of N. Fill in the missing values in the table.

N	Pattern of Rolls	Probability of N
1		1/6
2		
3	(Not 6, Not 6, 6)	
4		
5		
6		
7		
8		
9		
10		
11		
12		

Q.2 You have 10 keys on your key ring. All the keys look exactly alike. You arrive at your front door at midnight tonight. You try one of the keys in the door and if it opens you enter and place your keys back in your pocket. If however the door does not open you take the keys out of the door and they happen to fall on the ground. You pick up the keys not being able to identify which key you had previously used you try any one of the 10 keys in the door. You repeat this entire procedure, dropping the keys after each unsuccessful entry, until you eventually gain entry to your house.

i. What is the probability that you gain entry on the first attempt?

ii. What is the probability that it takes you three attempts to gain entry?

Each attempt at entry takes a total of 30 seconds to complete.

iii. What is the probability that you gain entry in 2 minutes exactly?

iv. What is the probability that you gain entry in a time not exceeding 3 minutes?

v. What is the maximum number of attempts it could take to enter your house?

vi. What is the probability that it takes at least 3 attempts to enter your house?

vii. If every night you and everyone else in the world were to repeat this procedure until the end of the world, and each time a record of the number of attempts that it took to enter houses was taken. What would be the average number of attempts?

Activity 7.6

Roll a fair die until the third 1 appears. Record the outcome of each roll on a piece of paper, classifying any roll which does not lead to a 1 as a failure and any roll leading to a 1 as a success. Now repeat this experiment 99 more times. For each of the experiments count the number of successes that occurred and the number of failures. Group the 100 experiments according to N, where N is the total number of rolls required to achieve the third success.

i. What is the minimum value for N in the 100 rolls?

ii. What was the largest value for N in the 100 rolls?

iii. What's the smallest possible theoretical value for N?

iv. What is the largest possible theoretical value for N?

v. For each value of N in your 100 experiments, write down a relative frequency for N.

You have now constructed a relative frequency table for N from the sample of experiments that you conducted. Can you construct a theoretical probability distribution for this experiment? For each value of N you will need to compute the probability of N occurring.

vi. What do you know about each experiment?

vii. What happens on the last roll?

viii. How many successes are achieved in the preceding N–1 rolls?

ix. What different possibilities are there for the way that those successes were arranged among the N–1 rolls?

x. What is the probability of getting a success on any individual roll?

xi. What is the probability of getting a failure on any individual roll?

xii. What is the probability of getting 3 successes AND N–3 failures?

xiii. How many different ways can 3 successes and N–3 failures occur in a set of N rolls, remembering that the game always ends when the third 1 is obtained?

Q.3 Ireland has made the final of the FIFA World Cup. The game has ended scoreless after extra time has been played and you are the only player on the Irish team willing or able to take penalty kicks. As the goalkeeper of the opposing team also plays for the same team as you in Serie A you have practised taking penalties against him many times and you know that you are successful exactly one quarter of all times you take a penalty against him.

i. What is the probability that you score on all of the first five kicks?
ii. What is the probability that your first successful kick occurs on the 4th attempt?

Suppose that under new rules the opposing team gets to take five penalties before you take any and the Irish goalkeeper saves all but one of the first 5 penalties that the opposing team takes. So this means you can win the World Cup if you score two penalties in the first 5 attempts.

iii. What is the probability that you will win the World Cup for Ireland with your first two kicks?
iv. If your winning kick occurs on the 3rd kick what are the possible occurrences on the 1st and 2nd kicks that you took?
v. What is the probability of each of these occurrences which led to you winning the World cup for Ireland on the 3rd kick?
vi. Are each of these possibilities that lead to you winning on the 3rd kick mutually exclusive or not?
vii. Considering the different things that might happen on your 1st and 2nd kicks, what is the overall probability that you will win the World cup on your 3rd kick?
viii. If you need two successful kicks to win, what is the probability that you win on the 5th and final kick?
ix. What is the probability that you win on the 4th kick?

As soon as you have scored two goals Ireland has won and you do not need to take any more penalties. What is the probability that you can win the World Cup for Ireland with the five kicks that are available to you?

The following activities can be conducted by groups of two people, one carrying out the activity and the other recording the results obtained. A minimum of ten groups is required.

Activity 7.7

Each of ten people should roll a fair die one hundred times, with the outcome of each roll recorded. Count the number of occurrences of each outcome for each individual and construct a relative frequency table and relative frequency histogram for this experiment.

a. What do you notice about the relative frequency of each of the outcomes 1, 2, 3, 4, 5 and 6?

Now combine the results of the experiments for all of the people, so that instead of having data for one hundred rolls you have data for at least one thousand rolls.

Again construct a relative frequency histogram for the experiment. What class intervals should you use for this histogram?

b. What do you now notice about the relative frequency of each of the outcomes 1, 2, 3, 4, 5 and 6?

Now using a calculator or a spreadsheet, compute an average (mean) of the outcomes that occurred on the first roll of the die for each person. Record this number and repeat this averaging process for each of the one hundred rolls of the die.

Now construct a relative frequency histogram using these one hundred averages.

c. What do you notice about the shape of this histogram? Is it the same as you had before you took averages?

Activity 7.8

Each of ten people should toss a coin until they get the first head, recording N, the number of tosses required until the first head was achieved. This is repeated a further ninety nine times by each person, to give a total of 1000 results. Construct a relative frequency histogram for N.

a. What does the histogram look like?
b. Are all of the bars of roughly equal height as they were when you rolled a die one hundred times?

Now combine all of the data from each person until you have results from at least one thousand tosses. Construct a relative frequency histogram for N using all of the data.

c. Has the shape of the histogram changed a lot from the ones constructed by each individual person?

Now average the values of N that were achieved in the first attempt at the game by each of the people participating. Remember you should have at least ten people playing this game so you should be averaging at least ten numbers. Now average the values of N from each person for the second game and repeat this averaging for each of the one hundred games until you have one hundred averaged values.

Construct a relative frequency histogram for these averages using class intervals of width 1.

d. What shape does this histogram have?

e. Is it similar to the histogram which was constructed prior to the averaging process?

f. Compare this histogram to that which was constructed by averaging the outcomes from the die rolling experiment. Do you notice any similarity?

Activity 7.9

Each of ten people should take a deck of cards and remove all of the picture cards and the jokers so that what remains are cards with denominations 1 to 10.

Each person shuffles their smaller pack of cards and randomly chooses one card from the deck, recording the card chosen on a piece of paper. The card is replaced in the deck and the cards are shuffled. Another card is chosen at random and its face value recorded; it is then replaced and the deck shuffled as before. This drawing of cards is repeated until one hundred cards have been drawn.

a. Construct a relative frequency histogram for this experiment using the data for each individual person.

b. Comment on the shape of this histogram.

c. Now amalgamate all of the data from all of the people playing the game and construct another relative frequency histogram.

d. Comment on the shape of this histogram.

Average the face values of the cards drawn by all the players on their first draw. Repeat this averaging process for each of the remaining ninety nine draws.

Using these one hundred averages construct a new relative frequency histogram with class interval widths equal to 1.

e. Comment on the shape of this histogram. Is it similar in shape to any other histograms you have constructed?

Activity 7.10

Open a blank worksheet in Excel and type =10*RAND() into cell A1.

Go to the Help Menu in Excel and search for help on the function RAND(). Having discovered what RAND() does can you explain =10*RAND() ?

Selecting the bottom right hand corner of Cell A1 with your mouse drag the formula down until it is entered into cells A1 to A400.

Select Column A with your mouse and paste a copy of it in each of columns B, C, D, E, F, G, H, I , J and K.

You will notice that the entries in each of the copies of Column A are not exactly the same. This is because the function RAND() updates itself each time it is copied.

Using the Help Menu in Excel search for help on the function COUNTIF.

In cell M1 enter =COUNTIF(A1:A400, '<=0.5')-COUNTIF(A1:A400,'<=0')

In cell M2 enter =COUNTIF(A1:A400, '<=1')-COUNTIF(A1:A400,'<=0.5')

Enter similar formulae, changing as appropriate for each of the cells M3 to M20 until you have entered =COUNTIF(A1:A400, '<=10')-COUNTIF(A1:A400,'<=9.5') in M20

a. Using the values in M1 to M20 construct a relative frequency histogram for the data in A1 to A400

What is the shape of the histogram?

In Cell L1 enter =AVERAGE(A1:J1)

By selecting the bottom right hand corner of Cell L1 drag this formula until it is entered in Cells L1 to L400

In cell N1 enter =COUNTIF(L1:L400, '<=0.5')-COUNTIF(L1:L400,'<=0')

In cell N2 enter =COUNTIF(L1:L400, '<=1')-COUNTIF(L1:L400,'<=0.5')

Enter similar formulae, changing as appropriate for each of the cells N3 to N20 until you have entered =COUNTIF(L1:L400, '<=10')-COUNTIF(L1:L400,'<=9.5') in N20

Using the values in N1 to N20 construct a relative frequency histogram for the data in L1 to L400

b. What is the shape of the histogram?
c. Is it similar to the histogram for the entries in column A?
d. Is it similar to any other histograms that you have produced?

Q.4 Continuous Distributions

A machine is created to select real numbers between 0 and 10. This machine is constructed to select all numbers in an equally likely manner so that no number is more likely to be chosen than any other number. The machine operates only once and produces one real number between 0 and 10.

a. What is the probability that it chooses the number 1.0786578?
b. What is the probability that it chooses the number 4.0?
c. What is the probability that it chooses a number which is greater than 0 and less than 5?
d. What is the probability that it chooses a number which is greater than or equal to 0 and less than or equal to 5?
e. If Z represents the number chosen by the machine, what is $P(1<Z<3)$?
f. What is $P(Z=2)$?

Q.5 Continuous Distributions

The author of this question has chosen a real number between 0 and 100 and has written this number down to its full accuracy of all of its decimal places on an extremely large piece of paper. The author will offer a substantial prize of one hundred million euro to anyone who guesses this number to all decimal places. After you have completed your Leaving Certificate you may enter this competition as many times as you like. Entry to the competition is entirely free.

a. Would it make sense to spend four years of your life attempting to guess this number rather than working or pursuing further study?
b. How many integers are there between 0 and 100?
c. If the author of this question had written an integer value between 0 and 100 down on a piece of paper and you were asked to guess an integer value between 0 and 100 what is the probability that you could guess the same value as the author chose?
d. If the author instead chose a real number between 0 and 1 would this game be easier to win?
e. How many possible real numbers are there between 0 and 1?
f. What is the probability that you could correctly guess the exact real number between 0 and 1 that the author wrote down?

Activity 7.11

Figure 1 contains a picture of the density function for a Standard Normal Random Variable Z.

a. The area under the curve between any two values, Z_1 and Z_2, on the x axis measures what quantity for the random variable Z?

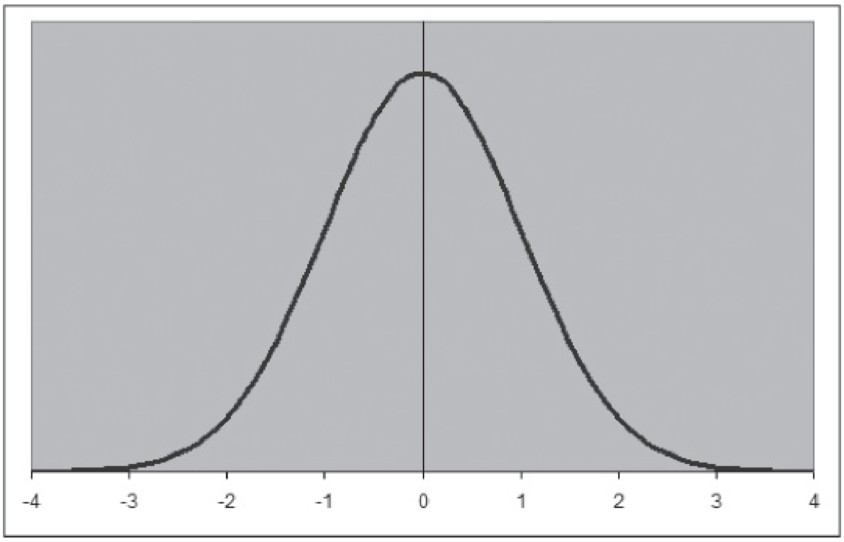


Figure 1 – The Standard Normal Density Function

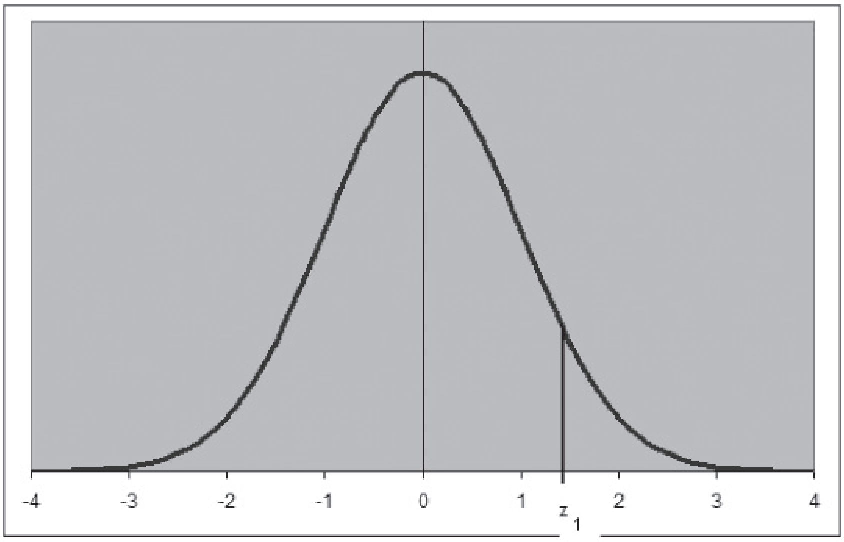


Figure 2: The Standard Normal Density Function

A number Z_1 is entered above on the horizontal axis of graph of the Standard Normal Density Function. Using a red pen enter the number '$-Z_1$' on the horizontal axis of this graph.

a. Shade in the region representing $P(Z > Z_1)$. Shade in the region representing $P(Z < -Z_1)$.

i. The area under the entire curve represents $P(-\infty < Z < +\infty)$. Considering this, what is the numerical value of the area under the entire curve?

ii. What is the relationship between $P(Z > Z_1)$ and $P(Z \le Z_1)$?

iii. What is $P(Z > 0)$? What is $P(Z < 0)$? What is $P(Z \le 0)$? What is $P(Z=0)$?

b. By referring to the Standard Normal Tables in your set of Mathematical Tables answer the following questions:

(i) What is $P(Z<1)$? (ii) What is $P(Z<2)$? (iii) What is $P(Z>1)$?
(iv) What is $P(0<Z<1)$? (v) What is $P(-1<Z<0)$? (vi) What is $P(-1<Z<1)$?
(vii) What is $P(-2<Z<2)$? (viii) What is $P(-3<Z<3)$?

In Figures 1 and 2 the density function is centred on the value 0. This is because the mean of Z is 0.

c. Draw a new graph, Figure3, which is identical in shape to that in Figure 1 but which is centred on 5 instead of on 0. Instead of Z we will name the variable whose density function is displayed in this new graph as W.

(i) What is the mean of W? (ii) What is E(W), the expected value of W?

Looking at the new graph that you have drawn, answer the following questions:
a) What is $P(-\infty < W < +\infty)$? b) What is $P(W>5)$? c) What is $P(W<5)$?
d) What is $P(W=5)$? e) What is the relationship between W and Z?

d. Suppose a new variable X is created where X=2Z. Answer the following questions.

(i) What is the mean of X? (ii) How does the mean of X compare with the mean of Z?
(iii) What is $P(-\infty < X < +\infty)$? (iv) What is $P(X>0)$? (v) What is $P(X<0)$? (vi) What is $P(X=0)$?

By referring to the Standard Normal Tables in your set of Mathematical Tables, and by remembering the relationship between X and Z answer the following questions:

(vii) What is $P(X<2)$? (viii) What is $P(X<4)$? (ix) What is $P(X>2)$? (x) What is $P(0<X<2)$?
(xi) What is $P(-2<X<0)$? (xii) What is $P(-2<X<2)$? (xiii) What is $P(-4<X<4)$?
(xiv) What is $P(-6<X<6)$?

e. Recall that the mean of a random variable determines where the density function is centred, and that the standard deviation of a random variable determines the width of the density function.

i. Compare $P(0<Z<1)$ with $P(0<X<2)$. The standard deviation of Z is 1, what is the standard deviation of X?

ii. Suppose H is a new random variable with mean 0 and standard deviation 3, what is $P(0<H<3)$?

iii. Suppose G is a random variable with mean equal to -2 and standard deviation equal to 1, draw a graph of the density function of G. a) What is $P(G> 0)$? b) What is $P(G>-2)$?

Q.6 Normal Distribution

X is a random variable with mean equal to 3 and standard deviation equal to 2.
(i) What is $P(X>4)$? (ii) What is $P(X>0)$? (iii) What is $P(X<-2)$? (iv) What is $P(0<X<3)$?

Q.7 Normal Distribution

Suppose the scores on the Leaving Certificate Mathematics Exam turn out to be normally distributed with a mean of 60% and a standard deviation of 15%.

i. What is the probability that a randomly selected student scores 75% or above?
ii. What is the probability that a randomly selected student scores 30% or below?

Q.8 Normal Distribution

i. The mean percentage achieved by students on a Statistics exam is 60%. The standard deviation of the exam marks is 10%. What is the probability that a randomly selected student scores a percentage above 80%? What is the probability that a randomly selected student scores a percentage below 45%? What is the probability that a randomly selected student scores a percentage between 50% and 75%?
ii. Suppose you were sitting this exam and you are offered a prize for getting a mark which was greater than 90% of all other students sitting the exam. What percentage would you need to get on the exam to win the prize?

Q.9 The heights of Irish women are normally distributed with mean 5 feet 6 inches and standard deviation 3 inches. Using this information, approximately what proportion of Irish women are taller than 6 feet? (Note there are twelve inches in one foot).

STATISTICS 1

SYLLABUS TOPIC: REPRESENTING DATA GRAPHICALLY AND NUMERICALLY

LEARNING OUTCOMES

As a result of completing the activities in this section you will be able to

- explore concepts that relate to ways of describing data, such as the shape of a distribution, what's typical in the data, measures of centre (mode, median, mean), and range or variability in the data
- use a variety of summary statistics to analyse the data: central tendency; mean, median, mode
- select appropriate graphical or numerical methods to describe the sample (univariate data only)
- evaluate the effectiveness of different displays in representing the findings of a statistical investigation conducted by others
- use pie charts , bar charts, line plots, histograms (equal intervals), stem and leaf plots to display data
- use back to back stem and leaf plots to compare data sets

There are links with Strand 3 (Number) where you will investigate models such as accumulating groups of equal size to make sense of the operation of multiplication.

INTRODUCTION

Being able to see a data set as a whole and so being able to use summary statistics such as averages to describe the 'big picture' or the overall shape of the data is an important learning intention of strand 1.

The activities described below allow you to investigate how the mean is constructed and the relationship of the mean to the data set it represents. You will also explore the different ways the median and mean represent the data - the median as a middle point in the data, and the mean as a 'point of balance' or the 'fair share' value of the data. Using two different representations of the mean gives you a chance to view the relationship between the mean and the data set through different models and so construct a firm understanding of the mathematical concept.

Prior learning

The idea that a set of data can be viewed and described as a unit is one of the key ideas about data that develops across primary school and is built on at second level. Initially, you looked at each individual piece of data. Gradually, you began to move away from a focus on individual pieces of data to looking at larger parts of the data. You learned to make general statements about the group of things or phenomena that the data represent, such as 'most people in our class have 1 or 2 siblings, and the range is from no siblings to 6 siblings.' Now you are ready to move away from making general statements and begin to make summary statements that describe the whole data set.

Activity 1.1

There are 5 bags of sweets, each of a different brand. All bags are the same size. The average price for a bag is €1.43

a. What could the individual prices of the 5 bags be? Think of at least two different sets of prices.

b. If both of your sets of prices included €1.43 as a price for at least one of the bags, price the five bags without using €1.43 as one of the prices.

c. Did you use €1.43 as the median? If so, what is the mean for your sets of prices? If you didn't use €1.43 as the median, what is the median for your sets of prices? Are the mean and median the same or different?

Discuss one of your lists of five prices with your group. How did you decide on your list of prices? How do you know what the average is in each example?
Note to each small group: Make sure you consider some lists that do not include a value of €1.43 as one of the prices.

d. There are seven bags of beads. Five of the bags have the following numbers of beads in them: 5, 7, 8, 9, and 12. Now work through parts (i), (ii) and (iii) with your group.

   i. Make a representation of the five bags by using small objects such as cubes, counters, marbles, etc. Make another representation of the five bags on a line plot.

   ii. Now use your representation to figure out how many beads could be in the other two bags so that 8 is the mean number of beads for all seven bags. Try to figure this out without adding up the beads in the five bags. Find at least two different sets of numbers for the two bags that will solve this problem.

   iii. Revise your two representations – counters and line plot – so that they show all 7 pieces of data. Can you 'see' the average in your representation?

e. What is the least number of beads there could be in one of the additional bags? What is the greatest number?

f. What numbers of beads could be in the two other bags if the mean number of beads was 7? What if the mean number was 10?

Q1. A teacher had some cards with groups of numbers displayed on them, as shown below

1,7,-8, 0,	0,0,0	-2,8,-6, 7, 11
0,11,8,0,13	-5,-4,-3,-2,-1 0,1,2,3,4,5	2,3,4,5,6,7,8 9,10

John was asked to calculate the mean of the numbers on each card and to put the cards that had a **mean of zero** into a box.

a. Circle the cards that John should put into the box.

The teacher has another card and tells the students that the mean of the numbers on this card is also zero.

b. Tick the correct box for each statement about this extra card.

Statement	Must be true	Could be true	Cannot be true
All of the numbers are *zero*			
Some of the numbers are *zero*			
There are as many negative numbers as positive numbers			
The sum of ***all*** the numbers is zero			
All of the numbers are positive numbers			
Some of the numbers are positive numbers			

Q.2 3 girls and 5 boys received text messages

The mean number of messages received by the 3 girls was **31**.

The mean number of messages received by the 5 boys was **27**.

Decide whether the following statements are true (T) or false (F), and justify your answer in each case:

i. The person who received the most messages must have been a girl.

ii. The mean number of messages for the 8 people was 29.

Q.3 Three girls and five boys were studying climate change in various countries around the world. They were examining the maximum daily temperatures in these areas

The mean daily temp of the locations studied by the 3 girls was 31° C
The mean daily temp of the locations studied by the 5 boys was 27 ° C

Decide whether the following statements are True or False, and justify your answer in each case.

i. The person who encountered the max daily temperature must have been a girl.
ii. The person who encountered the min daily temperature must have been a boy.
iii. The mean max daily temperature encountered by the 8 people was 29° C. .

Q.4 Sophie has six cards, each of which has a positive whole number printed on it. Four of the cards each have the number 9 on it.

a. Without knowing the numbers on the other two cards, can you give the value of the
   i. median
   ii. mode
   iii. range

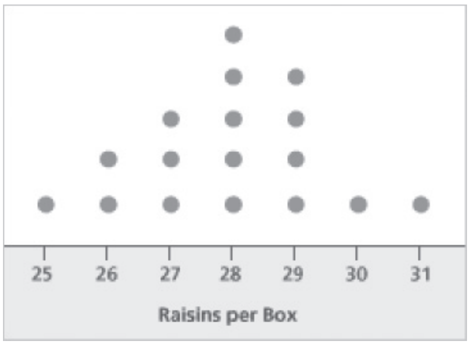
Explain your reasoning.

b. You are told that the six cards have a mean of 9. Give some possible whole numbers that could be on the other two cards. Which of your answers would give the greatest range? Why?

If the six cards have a mean of 9 and a range of 6 how many answers can you now find for the numbers on the remaining two cards?

Q.5 Students were investigating the number of raisins contained in individual mini-boxes of Sun-Maid raisins.

They recorded their results in the diagram shown.



a. Use the diagram to answer the following:

   i. How many boxes of raisins did they survey?

   ii. What was the modal number of raisins per box?

   iii. What is the median number of raisins per box? Explain how you found this answer.

b. If the students chose a box at random from all the boxes they surveyed what is the probability that the box contained 29 raisins?

Having done this activity, the students are asked to write down the answer they would give to the question: 'How many raisins are in a mini-box of Sun-Maid raisins?' Here are some of the answers they wrote down:

A 'There could be any number of raisins in a box.'
B 'There are about 28 raisins in a box.'
C 'There are almost always 28 raisins in a box.'
D 'You can be fairly sure there are 27, 28 or 29 raisins in a box.'
E 'Probably 28'.

c. Which of the answers above do you think is the best answer to the question? Explain why you think it's the best.

d. Which of the answers above do you think is the worst answer? Explain why you think it's the worst.

Extension (LC-OL)

Two extra boxes were found after the students had completed the diagram above. When the contents of these two boxes were added to the data, the mean number of raisins per box became 28.

e. Give one possible value each for the number of raisins in the two extra boxes.

Explain how you decided on these two numbers.

Extension (LC- HL)

The students wonder whether their sample was typical. On searching the internet, they read about a much larger-scale experiment on mini-boxes of Sun Maid raisins which indicates that the true mean is 31 raisins, with a standard deviation of 4.5. Some are surprised that the mean of the sample examined by the class was so much lower; others are not surprised, saying that this kind of variation between different samples is to be expected.

f. Calculate the probability that a random sample of 19 boxes will have a mean of less than or equal to 28 raisins per box.

g. What does this probability tell you about whether or not the sample examined by the class should be regarded as surprising?

Activity 1.2

A good part of one's day is spent travelling from one place to another. How much time do you spend travelling to school? How much time do your classmates spend travelling to school?

Carry out a survey to find out how everyone in your class travels to school, and how long the journey takes, on a given day. Your survey should enable you to answer a series of questions.

Deciding to walk or to go by car may depend on the distance, but, after choosing the method of transportation, does everybody spend about the same amount of time travelling to school?
Do those who take the bus to school spend less time than others?
Does the time it takes to get to school depend on where you live?

To better understand the situation, consider the 'time travelling to school' variable. Analyse the data you collect based on the method of transportation used.
Do you think this situation varies from one region in Ireland to another?

Time to get to school

Enter the class data in a table, such as the one below, grouping them in intervals of ten minutes, for example. First write down the numbers as you collect them. Then put them in ascending order to create a stem and leaf plot, where the tens are the 'stems' and the units are the 'leaves'. For example, a time of 15 minutes is recorded by placing a '5' in the Units column in the row which corresponds to the '1' in the Tens column.

Time to get to school Raw data	
Tens	Units
0	
1	
2	
3	...
...	...

Now, try to get an overview.

1. Look at all the ordered data. Half the class takes less than how many minutes to get to school? This number is called the median; it's the central value that divides the list of ordered data into two equal sections.
2. What is the average time that students in your class spend travelling to school?
3. Which row contains the most data? In your opinion, what does this mean?
4. What is the shortest time? What is the longest? What is the difference between them?
5. What can you say about the time that students in your class spend to get to school?

To get a better picture of the situation, it would help to add a column to your table that shows the number of students.

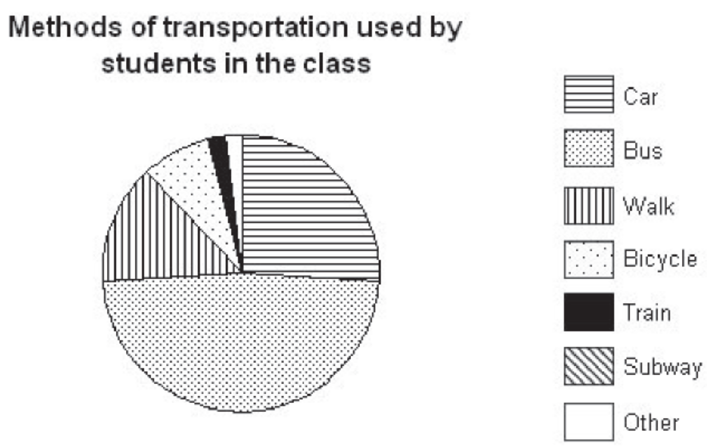
Time to get to school Raw data		
Tens	Units	No of Students
0		
1		
2		
3	...	
...	...	
	Total	

6. Now, can you create a graph that shows how much time the students in your class spend travelling to school? As you can see, everybody does not spend the same amount of time travelling to school.

You can now examine whether this time changes with the method of transportation.

Time spent by method of transportation

First, group together the students who use the same method of transportation. You can quickly determine the distribution of students by transportation method by creating a pie chart with a spreadsheet program. Your chart might look something like this:



From your chart, what are the most popular methods of transportation?
Approximately what fraction of the students in your class walk to school?

Now, for each method of transportation:

a. sort the time spent getting to school, from the shortest to the longest time.

b. determine the total time spent, which lets you calculate the average.

c. find the number of minutes or less that the faster half of the students spent travelling to school. This is the median or the value of the middle item of the ordered data.

d. add the minimum and maximum amount of time spent travelling to school.

Create a descriptive table that will look like this:

Method of Transportation	Time to get to school (mins)	No	Total Time	Average Time	Median	Min	Max
Car	5, 12, 12, 2, 32,	5	83	83/5	12	5	32

You can now examine the time by method of transportation.
Do you notice any significant differences?
Which method of transportation takes the longest?
Which method of transportation shows the biggest difference between the shortest time (minimum) and the longest time (maximum)? What might explain this?
Can you describe the overall situation for your class and present your point of view? What type of transportation do you think we should encourage? Under what conditions? Why?
Finally, use the data you have obtained to create a graph that properly conveys the information about your class that you feel is important.

Comparing your class to a sample of Irish students

Do you think the situation of your class resembles that of most Irish students?
Obtain a sample of 50 students from your school. Then do the same analysis that you did for your own class.
Is the time spent getting to school approximately the same for both groups? If not, how does it vary?
To help you better compare the data, create two tables side-by-side for each group.

Time to get to school				
Raw data for the school			Raw data for our class	
Students	Units	Tens	Units	Students
		0		
		1		
		2		
		3	...	...
		4	...	...
		5	...	...
50	Total	TOTAL	Total	

'A picture is worth a thousand words' and can certainly make it easier to read all these numbers. Create appropriate graphs to easily compare the time spent getting to school for both groups.

You can also compare the methods of transportation used.
For each group: create a pie chart to illustrate the distribution of students for the different methods of transportation used to get to school.
Use a descriptive table to examine the time spent by method of transportation used.
Do you arrive at the same observations for both groups? Are there any significant differences? If yes, what are they? Can you explain the differences taking into account the characteristics of your region?
Create a visual representation that properly illustrates and conveys your main conclusions.

STATISTICS 2

SYLLABUS TOPIC: FINDING, COLLECTING AND ORGANISING DATA

LEARNING OUTCOMES

As a result of completing the activities in this section you will be able to

- clarify the problem at hand
- formulate one (or more) questions that can be answered with data
- explore different ways of collecting data
- design a plan and collect data on the basis of above knowledge
- generate data, or source data from other sources including the internet
- discuss different types of studies: sample surveys, observational studies and designed experiments
- select a sample (Simple Random Sample)
- recognise the importance of representativeness so as to avoid biased samples
- design a plan and collect data on basis of above knowledge.

The activities described below and the questions that follow give you the opportunity to construct an understanding of the concept of finding, collecting and organising data in a statistical investigation. By carrying out a complete data investigation, from formulating a question through drawing conclusions from your data, you will gain an understanding of data analysis as a tool for learning about the world.

The activities are designed to build on your previous experiences with data, and to introduce you to the ideas you will work on as you progress through statistics in Strand 1.

During these activities you will work with categorical data, noticing how these data can be organised in different ways to give different views of the data.
As a result you should be able to

- gather data from a group
- classify the data
- write sentences that describe the 'Big Picture' of the data
- appreciate how the purpose of the research will affect how the data is gathered
- understand that the way data is represented can illuminate different aspects of the data.

Activity 2.1: A data Investigation

With what well-known person would you like to meet?

1. You will be working in groups on a data investigation. The first step is for each student to decide on his/her own how they would answer the survey question. Each student will need to write their answer a number of times on separate pieces of paper so that they can give their individual answers to each group, including their own.
2. Each group collects answers from everyone; make sure your group has a full class set of data that you can discuss.
3. Before you look at the data spend a few minutes discussing what might be interesting about them.
4. As a group sort the class data into three piles according to what they have in common. This is called classifying your data.
5. Choose one of your ideas for sorting and arrange your cards on a large piece of paper to show that classification
6. Write a sentence or two on your display that tells what you notice about the data
7. Post your display on the wall. If you finish before other groups, discuss issues about data that arose while you did this activity.
8. Can you represent this data in a chart?

Key Words: **Category, Data**

As you work through this activity reflect with your group on

- What issues came up for you as you tried to represent these data?
- What does the data tell you about the group?
- What questions arise for you while looking at this data? How might you modify the survey in order to address these?
- Did everyone interpret the original question in the same way?
- What were you thinking when you made your own decision?

Consider the following question:

How many countries have you visited?

Elect a scribe to sketch a line plot with reasonable intervals on the board. Collect data on the line plot by marking an X for the value of each person's response. (Note: a line plot is a graph for numerical data that is similar to a bar chart. It is one of the plots in common use in statistics.) Try to form statements that describe the data. What can they say for the class as a whole about the number of countries that they have visited?

Activity 2.2

1. Note: You have 30 mins to complete this assignment and post a representation of your data for others to see. That means you will need to decide on a question and collect your data efficiently. You may need to design a data collection sheet. Think about how you will make sure you get a response from every person. After 15 mins you should be ready to start making a data representation. Your representation need not be decorative or elaborative. Focus on how well it communicates information about your data.
2. Select a question that will result in numerical data
3. Collect data from everyone in the class.
4. Create a line plot for your data
5. Write three to five sentences on your display that describe your data
6. When your display is complete, discuss issues that arose in your group as you defined your question
7. What further questions might you want to pursue based on these initial data?

Sample data collection sheet

Name	Name	Name	Name	Name
Name	Name	Name	Name	Name
Name	Name	Name	Name	Name
Name	Name	Name	Name	Name
Name	Name	Name	Name	Name
Name	Name	Name	Name	Name

STATISTICS 3

SYLLABUS TOPIC: REPRESENTING DATA GRAPHICALLY AND NUMERICALLY

LEARNING OUTCOME

As a result of completing the activities in this section you will be able to

- explore the distribution of data, including concepts of symmetry and skewness
- interpret a histogram in terms of distribution of data
- recognise standard deviation as a measure of variability
- make decisions based on the empirical rule

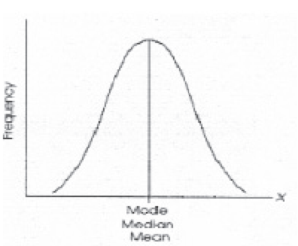
INTRODUCTION

The following activities and questions will enable you to construct an understanding of distributions and how the shape of the distribution relates to the mean, mode and median of the data. You will also construct an understanding of the concept of variability and how standard deviation relates to the representation of the data in a histogram. You will derive the empirical rule and make decisions based on it.

Activity 3.1

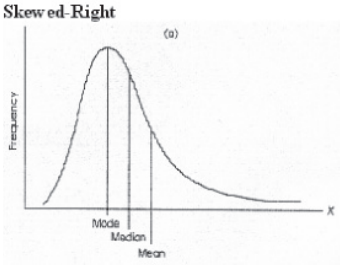
Examine the following distributions; note how the mean, median and mode compare in each situation and the shape characteristics of each distribution. A normal distribution is an example of a symmetric distribution.

Symmetric



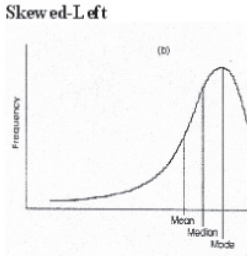
Mean= Median= Mode

Skewed-Right



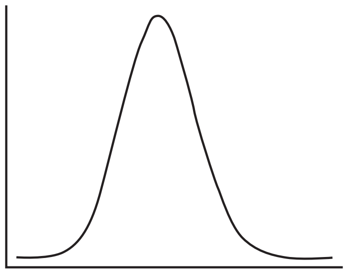
Mean> Median> Mode

Skewed-Left

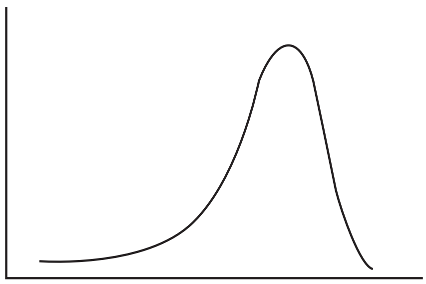


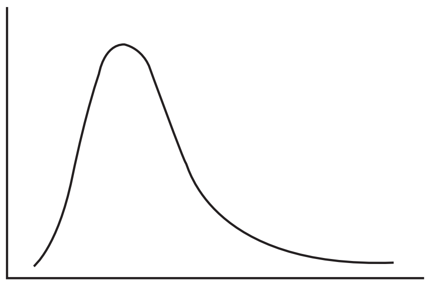
Mean< Median< Mode

Q.1 Examine the distributions sketched below. Label them as symmetric/normal, skewed right, or skewed left . Develop a list of situations in which the data gathered would produce each of the 3 different distributions.



……………………………………..





Q.2 Using the mean, median and mode, sketch the shape of the frequency histogram with the following characteristics

a. mean: 7.5 median: 6 mode: 5.7

b. mean: 6 median: 6 mode: 10, 12

c. mean: 7.5 median: 8.5 mode: 9

d. mean: 7.5 median: 7.5 mode: 7.5

Q.3 A pair of dice is rolled numerous times. The sum of the dice, as well as the frequency, is recorded. Calculate the mean, median and mode. Use these results to identify the shape of this distribution.

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	2	3	5	7	9	11	8	7	4	2	1

Q.4 The table shows information about the age of a sample of internet Facebook users

Age (t years)	Frequency
$10 < t \leq 15$	24
$15 < t \leq 20$	37
$20 < t \leq 25$	42
$25 < t \leq 30$	65
$30 < t \leq 35$	24
$35 < t \leq 40$	17

Draw a histogram of the data and calculate the mean and median age of the Facebook users.
How are the data distributed? Explain your reasoning

Q.5 A dataset consists of the prices of all new cars sold in Ireland last year. For this dataset which do you think would be higher, the mean or the median, or are they about equal?

Q.6 Draw a rough histogram of a dataset that is skewed to the right.

A dataset consists of the salaries in a company employing 100 factory workers and two highly paid executives. For this dataset which is higher, the mean or median or are they about equal?

Would the range or quartiles be more heavily influenced by outliers?

Q.7 The following data set represents the ages, to the nearest year, of 27 university students in a statistics class.

17 21 23 19 27 18 20 21 28 31
18 21 24 30 25 19 22 27 35 18
29 22 20 30 28 21 23

e. Determine the mean, median and mode for the data set

f. Do you think the data is normally distributed? give a reason for your answer

g. Determine the standard deviation of the data.

Q.8 A dataset consists of the ages when students in your class will get married. If you were to draw a histogram representing the relative frequency of marriage at different ages for the students in your class what would the histogram look like?

i. Would it be symmetric, skewed to the right or skewed to the left?

ii. For this dataset which do you think would be higher, the mean or median, or are they about equal?

iii. 'Most Irish people have more than the mean number of legs for human beings.' Explain how this can be, using the concepts of mean, median and mode.

iv. Explain how most Irish houses cost less than the average house price in Ireland.

Q.9 In a data set would the range, the standard deviation or the inter-quartile range be more heavily influenced by outliers?

A dataset consists of percentage scores obtained by students on an examination. The median of the students marks is 50%, the lower quartile (ie 25th percentile) is 25% and the mean of the marks is 70%. Would you expect the upper-quartile of the data set (ie the 75th percentile) to be greater or less than 75%?

What Makes the Standard Deviation larger or smaller?

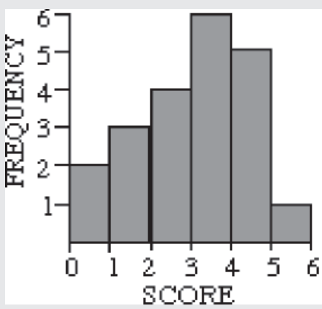
Study the 15 pairs of graphs that follow. The mean for each graph (μ) is given just above each histogram.

For each pair of graphs presented below:

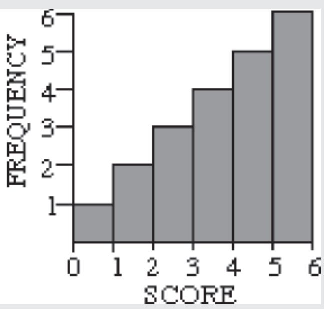
1. Indicate whether one of the graphs has a larger standard deviation than the other or if the two graphs have the same standard deviation.
2. Identify the characteristics of the graphs that make the standard deviation larger or smaller.

1.

A $\mu = 2.57$



B $\mu = 3.33$



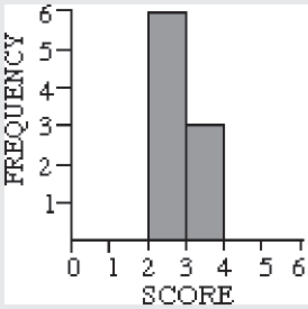
A has a larger standard deviation than B ☐

B has a larger standard deviation than A ☐

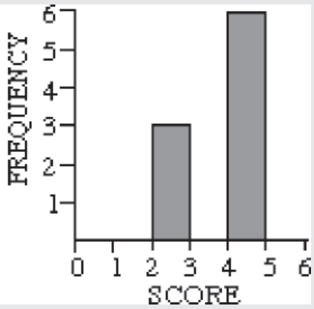
Both graphs have the same standard deviation ☐

2.

A $\mu = 2.57$



B $\mu = 3.33$



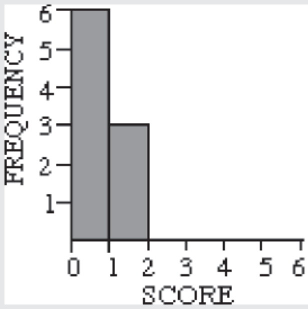
A has a larger standard deviation than B

B has a larger standard deviation than A

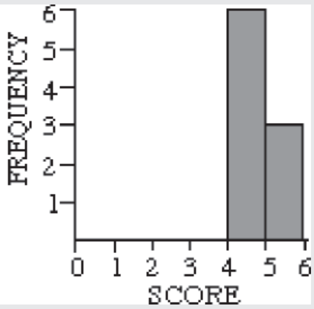
Both graphs have the same standard deviation

3.

A $\mu = .33$



B $\mu = 4.33$



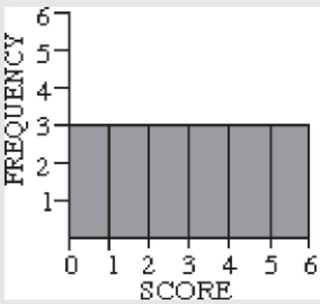
A has a larger standard deviation than B

B has a larger standard deviation than A

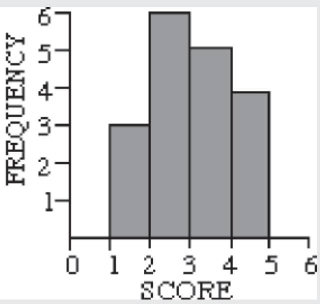
Both graphs have the same standard deviation

4.

A $\mu = 2.50$



B $\mu = 2.56$



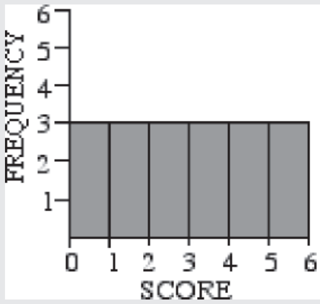
A has a larger standard deviation than B

B has a larger standard deviation than A

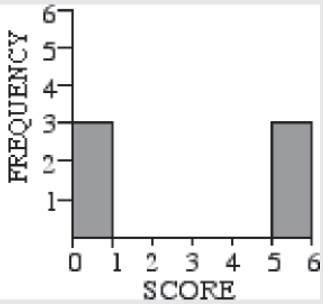
Both graphs have the same standard deviation

5.

A $\mu = 2.50$



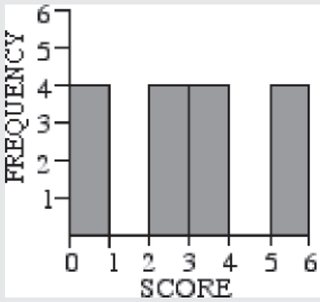
B $\mu = 2.50$



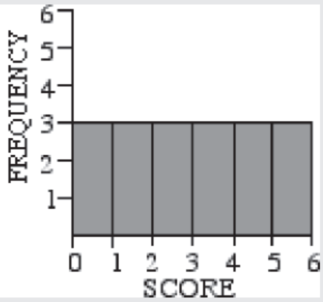
- A has a larger standard deviation than B ☐
- B has a larger standard deviation than A ☐
- Both graphs have the same standard deviation ☐

6.

A $\mu = 2.50$



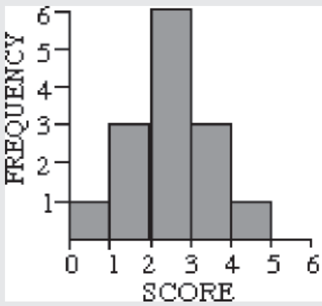
B $\mu = 2.50$



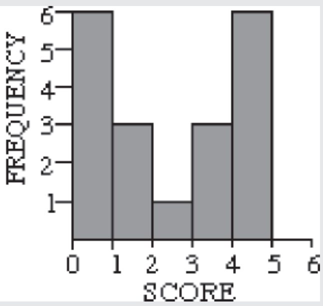
- A has a larger standard deviation than B ☐
- B has a larger standard deviation than A ☐
- Both graphs have the same standard deviation ☐

7.

A $\mu = 2.00$



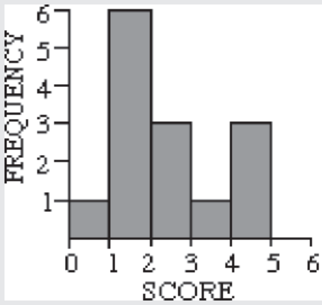
B $\mu = 2.00$



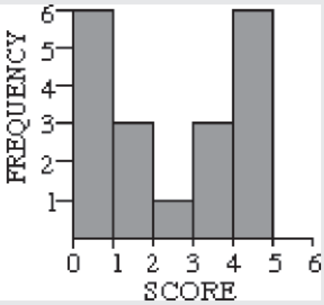
- A has a larger standard deviation than B ☐
- B has a larger standard deviation than A ☐
- Both graphs have the same standard deviation ☐

8.

A $\mu = 1.93$



B $\mu = 2.00$



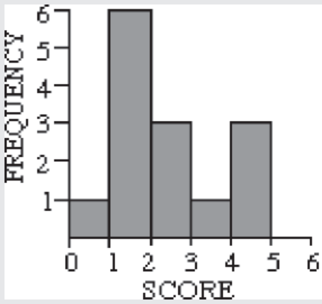
A has a larger standard deviation than B ☐

B has a larger standard deviation than A ☐

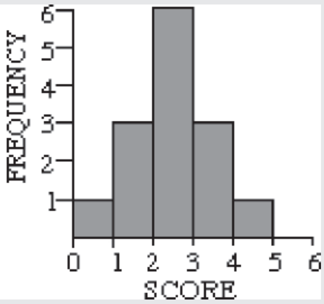
Both graphs have the same standard deviation ☐

9.

A $\mu = 1.93$



B $\mu = 2.00$



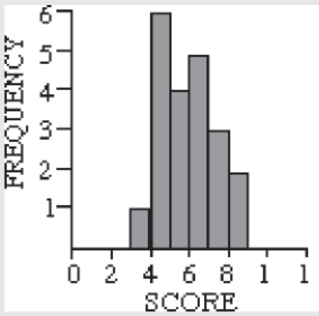
A has a larger standard deviation than B ☐

B has a larger standard deviation than A ☐

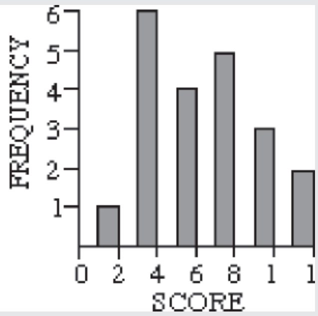
Both graphs have the same standard deviation ☐

10.

A $\mu = 5.43$



B $\mu = 5.86$



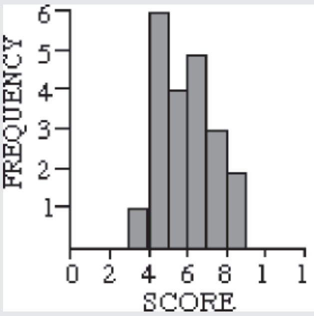
A has a larger standard deviation than B ☐

B has a larger standard deviation than A ☐

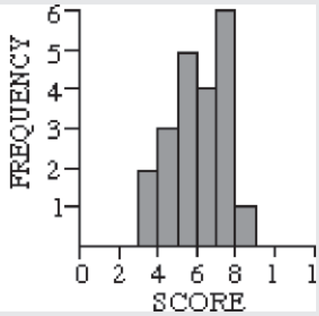
Both graphs have the same standard deviation ☐

11.

A μ = 5.43



B μ = 5.57



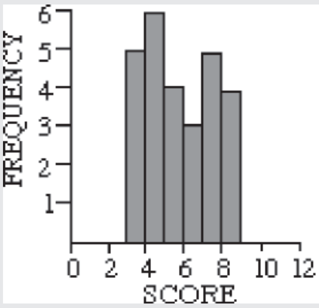
A has a larger standard deviation than B ☐

B has a larger standard deviation than A ☐

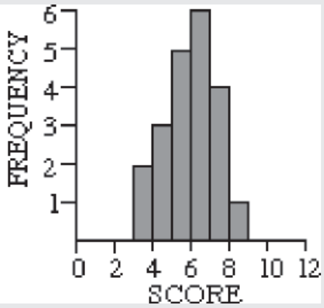
Both graphs have the same standard deviation ☐

12.

A μ = 5.33



B μ = 5.48



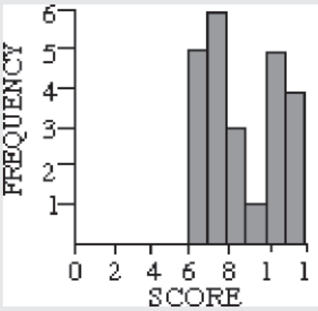
A has a larger standard deviation than B ☐

B has a larger standard deviation than A ☐

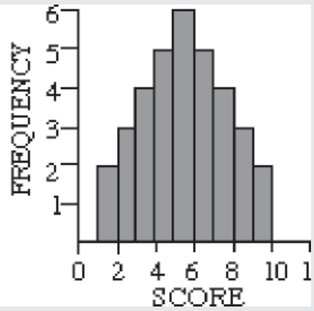
Both graphs have the same standard deviation ☐

13.

A μ = 8.33



B μ = 5.00



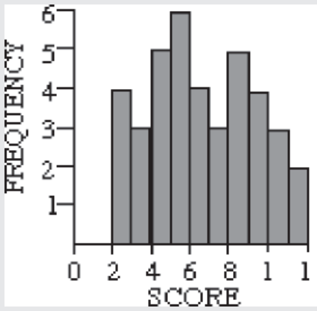
A has a larger standard deviation than B ☐

B has a larger standard deviation than A ☐

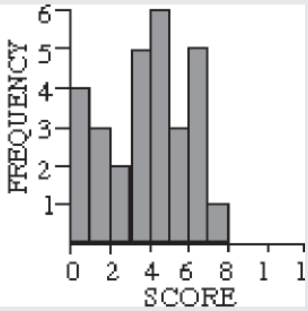
Both graphs have the same standard deviation ☐

14.

A $\mu = 6.15$



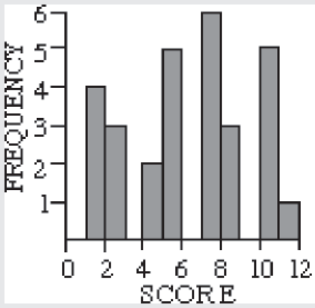
B $\mu = 3.38$



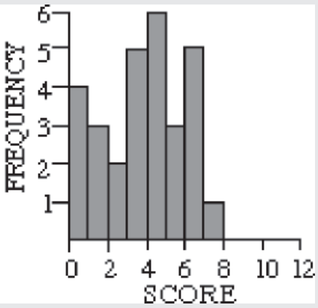
- A has a larger standard deviation than B ☐
- B has a larger standard deviation than A ☐
- Both graphs have the same standard deviation ☐

15.

A $\mu = 5.86$



B $\mu = 3.38$



- A has a larger standard deviation than B ☐
- B has a larger standard deviation than A ☐
- Both graphs have the same standard deviation ☐

Activity 3.2

The office manager of a small office wants to get an idea of the number of phone calls made by the people working in the office during a typical day in one week in June.

The number of calls on each day of the (5-day) week is recorded. They are as follows:
Monday:15; Tuesday: 23; Wednesday: 19; Thursday: 31; Friday: 22

1. Calculate the mean number of phone calls made

2. Calculate the standard deviation (correct to 1 decimal place).

3. Calculate 1 Standard Deviation from the mean:

$\bar{x} \pm \mathbf{s}$ = ...or ....

The interval of values is ($\bar{x} - \mathbf{s}$; $\bar{x} + \mathbf{s}$) = ...

4. On how many days is the number of calls within one Standard Deviation of the mean?

Number of days =
Percentage of days =
Therefore the phone calls on ....% of the days lies within 1 Standard Deviation of the mean.

Activity 3.3

The pilot study for a Census@School project gave the following data for the heights of 7 067 students from Grade 3 to Grade 11.

Height Less than (m)	Total number of students
$0 \le h < 1.06$	8
$1.06 \le h < 1.21$	111
$1.21 \le h \le 1.36$	1,114
$1.36 \le h \le 1.52$	2,218
$1.52 \le h \le 1.67$	2,413
$1.67 \le h \le 1.83$	1,105
$1.83 \le h \le 1.98$	108

The mean x = 1.52 cm and the Standard Deviation s = 15 cm

Draw a histogram of this data .Does the distribution look symmetric?

a.

i. Calculate $\bar{x} + s$

ii. Calculate $\bar{x} - s$

iii. What is the total number of students whose heights are less than $\bar{x} + s$?

iv. What is the total number of students whose heights are less than $\bar{x} - s$?

v. Calculate the number of students whose heights are within 1 standard deviation of the mean

vi. Write this number as a percentage of the total number of students

b.

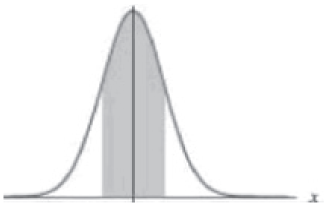
i. Calculate the number of students whose heights are within 2 standard deviations of the mean i.e. within the interval ($\bar{x} - 2s$; $\bar{x} + 2s$)

ii. Write this as a percentage

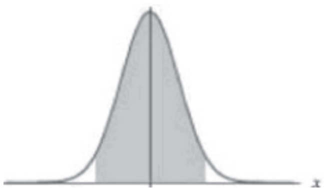
iii. Calculate the number of students whose heights are within 3 standard deviations of the mean i.e. within the interval ($\bar{x} -3s$; $\bar{x} +3s$)

iv. Write this as a percentage.

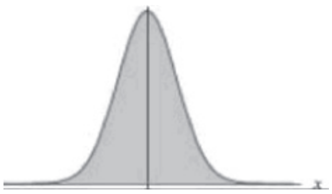
In the activities above you worked out the percentage of data items within 1, 2 and 3 standard deviations of the mean. Summarise your findings below. [Your findings are usually referred to as the empirical rule]



……% of the population falls within 1 standard deviation of the mean.



…….% of the population falls within 2 standard deviations of the mean.

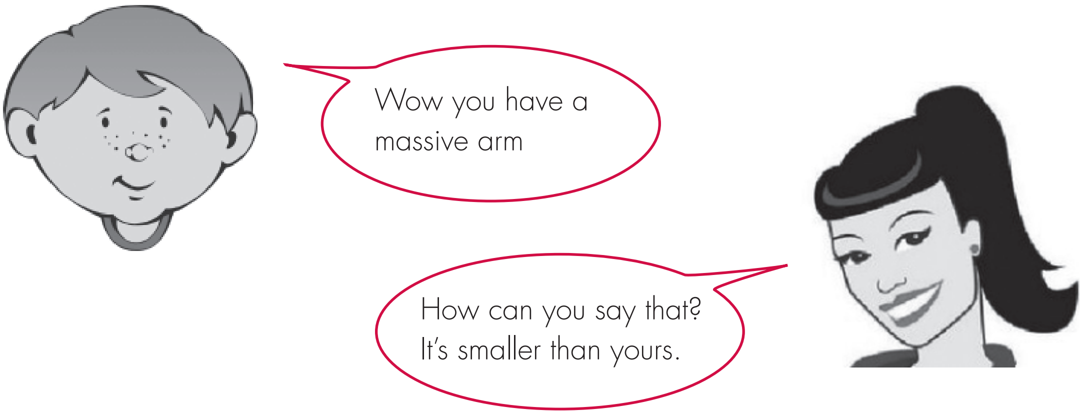


……% of the population falls within 3 standard deviations of the mean.

Q.10 For data that is symmetrically distributed and bell shaped (similar to a normal distribution), approximately what proportion of observations lies within one standard deviation of the mean according to the empirical rule? What proportion lies within two standard deviations of the mean? What proportion lies within three standard deviations of the mean?

Activity 3.4

Sophie and Jack were discussing proportions in an Art class. Sophie measured the length of her arm and found it to be 783mm long. Jack measured the length of his arm and found it to be 789mm long.



Later in the day their mathematics teacher suggested they use mathematics to see if there was any truth in Jack's statement. Sophie and Jack trawled the internet for data on arm lengths and found the results of a survey, which are summarised in the table below, recording the lengths of the arms of 500 females over the age of 16 and 500 males over the age of 16 measured from the shoulder to the finger tip with the arm outstretched.

Arm length (mm)	No of females	No of males	No of adults
$620 \leq x \leq 640$	3	0	
$640 \leq x \leq 660$	11	0	
$660 \leq x \leq 680$	41	0	
$680 \leq x \leq 700$	92	0	
$700 \leq x \leq 720$	132	2	
$720 \leq x \leq 740$	120	9	
$740 \leq x \leq 760$	69	27	
$760 \leq x \leq 780$	25	71	
$780 \leq x \leq 800$	6	114	
$800 \leq x \leq 820$	1	122	
$820 \leq x \leq 840$	0	89	
$840 \leq x \leq 860$	0	46	
$860 \leq x \leq 880$	0	15	
$880 \leq x \leq 900$	0	4	
$900 \leq x \leq 920$	0	1	
Total	500	500	

Work in groups of three to complete the following tasks.

1. Choose a data set to work with; female, male or adult.
2. Work with your set of grouped data and calculate
   a. The mean
   b. The median
   c. The standard deviation
3. Draw a histogram to illustrate your data
4. Check the following:
   a. Do the mean and median of your set of data have approximately the same value?
   b. Does approximately 99.7% of the data lie within three standard deviations of the mean?.
5. Compare the histograms and comment on similarities and differences. Which set of data, if any, is normally distributed?

Extension for HL

Refer to Jack's statement. Is Sophie's arm length unusual? Explain your answer.
What about Sophie's reply? Is Jack's arm length unusual? Explain you answer.

Is it reasonable to conclude from the data that men have longer arms than women?

Q.11 A recent survey of Irish school-going teenagers reported their 'Attitude-Toward-Authority' scores to have mean 107 and standard deviation 14 among the males, and 115 and 13, respectively, among the females. A score which is higher than 90 indicates pro-authority feelings.

1. Relative to his/her own group, who is more pro-authority: a male teenager with a score of 120 or a female with a score of 125?
2. Assuming that the scores are normally distributed, what proportion of the male teenagers can be considered pro-authority?
3. In a group of 250 female teenagers, how many do you expect to be pro-authority?
4. Ninety percent of female teenagers have their scores between 93.55 and what other score?
5. A teenager is considered rebellious if he is in the first percentile among his peers. Suppose Luke's score is 79; can we consider him rebellious?
6. Where does the Empirical rule say that 95% of the observations lie in a distribution which is approximately bell shaped?

Q.12 Record the heights in centimetres and weights in kilograms of all of the people in your mathematics class. Compute the mean and standard deviation of these heights and weights. How many heights in your data set were within one standard deviation of the mean height (i.e. between the mean minus one standard deviation and the mean plus one standard deviation)? How many were further than three standard deviations from the mean? Answer the same two questions for the weights.

Q.13 100 students sit a statistics examination and the marks scored by the students are found to be approximately normally distributed. The mean mark scored by students is 50% and the standard deviation of the students' marks is 5%. Two students scored 99%. These students were called out of class by the principal, who had a degree in statistics, and accused of cheating on the examination. Why did the principal feel that the students had cheated?

Q.14 Using the webpage http://uk.finance.yahoo.com/, or another source search, for historical share prices for any company listed there. From this webpage it is possible to download share price information into a spreadsheet. Download the daily share prices of any company for the last year.

a. For these share prices compute the mean, the median and the mode.

b. Compute the interquartile range, the range and the standard deviation.

c. Do any prices appear to lie more than one and half interquartile ranges above the upper quartile? Do any share prices lie more than three interquartile ranges above the upper quartile?

d. Do any share prices lie more than four standard deviations above the mean?

e. Construct a relative frequency histogram for the daily share prices. Do the share prices seem to be normally distributed?

STATISTICS 3

SYLLABUS TOPIC: FINDING, COLLECTING AND ORGANISING DATA

LEARNING OUTCOME

As a result of completing the activities in this section you will be able to

- clarify the problem at hand
- formulate one (or more) questions that can be answered with data
- explore different ways of collecting data
- design a plan and collect data on the basis of above knowledge
- generate data, or source data from other sources including the internet

Students working at OL will be given the opportunity to

- discuss different types of studies: sample surveys, observational studies and designed experiments
- select a sample (Simple Random Sample)
- recognise the importance of representativeness so as to avoid biased samples
- design a plan and collect data on basis of above knowledge

Whilst those working at HL will learn to

- recognise the importance of randomisation and the role of the control group in studies
- recognise biases, limitations and ethical issues of each type of study
- select a sample (stratified, cluster, quota etc. – no formulae required, just definitions of these)
- design a plan and collect data on basis of above knowledge

Activity 3.1

There are many situations where it is impossible or impractical to gather information on all the items in a survey.
List some such situations.

Consider the following

Frank has to find out about the breakfast eating habits of students in his school. The school is co-educational. There are 150 students in Transition Year (TY) and about 200 students in each of the other year groups.
Frank decides it is impractical to interview every student in the school, so he decides to gather the information from a sample of 100–200 students. Here are some ways for him to choose his sample

A. Frank can get to school early and use the first 100 students who arrive

a. Is any group of students likely to be under-represented?

b. Are equal numbers of boys and girls likely to be included in this sample?

c. Is age group likely to be represented in the sample in the same proportion as they are in the school?

d. Do you think a student's early arrival has any relationship with the breakfast they eat?

e. From the experiences of your own school, can you think of any reasons why this would not be a representative sample?

B. Frank can take one class from each year group as his sample.

a. If the classes are banded (i.e. grouped according to ability), is this likely to affect their breakfast eating habits?

b. Does the inclusion of a complete TY class reflect the proportion of these students in the whole school?

c. Are the age groups in the sample represented in the same proportion as they occur in the school?

d. Can you suggest ways of improving the selection of the sample to make it as representative as possible?

This kind of sample (identifying groups with different characteristics and taking a sample of each group) is called a stratified sample.

C. Frank can draw his sample from the main school register. The students are listed in this register with the boys first in alphabetical order followed by girls, also in alphabetical order.

Frank can take the first 50 boys and the first 50 girls as his sample. Alternatively, Frank can take every fifth student on the register as his sample. (This method of sampling is an attempt at random selection.)

a. Can you suggest reasons why either method may not give a sample that represents the whole school population?

b. Devise a way of selecting about 100 students from your own school so that they reflect, as nearly as possible, known characteristics of the whole set of students.

Construct a mind map or graphic organiser (see www.ncca.ie) to help you remember how to choose a representative sample from a population.

Q.1 Are men or women more adept at remembering where they leave misplaced items (like car keys)? According to University College Dublin researchers, women show greater competence in actually finding these objects. Approximately 300 men and women from Dublin participated in a study in which each person placed 20 common objects in a 12-room 'virtual' house represented on a computer screen. Thirty minutes later, the subjects were asked to recall where they put each of the objects. For each object, a recall variable was measured as 'yes' or 'no.'

a. Identify the population of interest to the researcher.
b. Identify the sample.
c. Does the study involve descriptive or inferential statistics? Explain.
d. Are the variables measured in the study quantitative or qualitative?

Q.2 Which of the following statements is correct regarding observational studies?

a. A researcher can observe but not control the explanatory variables.
b. A researcher can define but not observe the explanatory variables.
c. A researcher can minimise but not eliminate the explanatory variables.
d. A researcher can control but not observe the explanatory variables.

Q.3 Suppose you wanted to find out whether there should be fewer days in the school year. Select twenty people in your school and ask them whether schools should be closed for all of the months of May, June, July and August. What proportion of people agreed that the school should be closed?

If the Government were to hold a referendum tomorrow to decide whether to close schools for four months during the summer do you think that the same result would occur in the referendum vote as you got in your survey? Can you identify some reasons why your survey might disagree with the referendum vote?

Q.4 Many surveys today are conducted using the internet. A person, while browsing a particular website, is asked to participate in a brief survey. Do you think the results of such surveys would give the same answers as a survey of the entire population?

Q.5 Another method that survey companies have of selecting people is to hire people to go onto the street with a clipboard, stop people, and ask them questions. Do you think that the people who participate in these surveys are representative of the entire population? Can you identify any people who may not participate in a street survey like this? Would it make a difference if the survey was conducted on a Thursday in April at 2.00pm or the following Saturday at 3.00pm? Would students from secondary or primary schools be more likely to be included in one of these surveys? What about people who are at work?

Q.6 Some companies use computers to randomly dial phone numbers and, if someone answers the phone, a company employee asks the individual who answered the phone if they would participate in a survey. Do you think that a survey which is conducted like this gives an accurate reflection of the opinions of the entire population?

Q.7 A crèche needs to paint its building but has only two colours of paint available, pink and blue. The owners of the crèche will make the decision on which of the two colours to use by asking ten children which colour they should paint the building.

What do you think the answer will be if the ten children who are asked are all girls?
Do you think a different answer would be given if the ten children are boys?
Suppose there are fifty children in the crèche. Can you recommend a way to identify what proportion of girls and boys should be in the group of ten that are surveyed?

Q.8 Suppose that the government of Ruritania was concerned that mobile phones caused brain tumours. In order to determine if this is true they decided in 1980 to conduct a study in which they required one thousand people to speak on their mobile phones for ten hours every day. The study was conducted over twenty nine years and at the end of the study the number of people who developed brain tumours was measured in the group. This information was then be used to decide if mobile phones should be banned in Ruritania.

Comment on the design of this study. Do you think that the results of this study are valid? Do you think the study is ethical? How would you recommend a way to decide if mobile phones cause brain tumours?

Q.9 A company claims that it has developed a new drug which is completely harmless and has the ability to cure obesity. The company wants to market this new drug and is asked to prove that the drug works. The company says that it has conducted a 'clinical trial' of the drug in which one hundred people participated. There were two groups of fifty people in the trial. The treatment group was given the drug to take daily over a two year period and the control group did not take the drug. The company was able to show that the group who took the drug had an average weight of 50kg and the group who did not take the drug had an average weight of 80kg.

Based on this information do you think the drug works?

Suppose that you knew that the drug company had selected the fifty people for the treatment group from among the members of a gym and they chose the control group from those who ate everyday in a nearby fast food restaurant. Does this information change your opinion about the efficacy of the drug?

Suppose instead that the members of the groups were evenly balanced as to members of the gym and those who ate in the fast food restaurant. But it was noted that all of the members of the treatment group happened to be called Pauline and all the members of the control group were called Paul. Do you believe the drug was effective in reducing weight if this was the case?

When a company is selecting a control group and a treatment group to test a new drug, what differences or similarities should there be between the two groups?

In Sweden, many studies are conducted using pairs of twins. Why is this done? How would you recommend designing a study in which you had fifty pairs of twins available to you?

Q.10 Suppose that a tobacco company claims that they have discovered that one of their brands of cigarette does not cause cancer. The cigarette has been on the market for the last forty years. How should the claim that it does not cause cancer be tested? What is the best study design to use in this situation? What ethical issues may there be in designing this type of study? Is it possible to design a study to test this claim which does not pose any ethical problems?

Q.11 A teacher wants to determine which of her thirty students is best at solving a particular type of mathematical problem. To answer this question she has constructed thirty different examples of this problem and she assigns one problem to each student. Ten of the students must answer the problem on the blackboard in front of the entire class. Ten of the students must answer the problem in class while the teacher moves around behind their desks and looks over their shoulders as they solve the problem. The last ten students are allowed to answer the question in class at a time when the other twenty students are reading books and the teacher is at the top of the class also reading.

a. Do you think there will be a difference in the average performance of the three groups? Which group do you think will perform best?

b. Do you think that participants in a study perform differently depending on whether they are being observed or not?

c. Suppose that during one half of a double maths class period a teacher is present in class but during the following period the teacher is called away and asks the students to work on their own. During which period do you think students will work hardest?

Q.12 A survey is conducted in a school in the centre of Dublin to determine what students do in their time outside school. Do you think that the results of this survey would match the results of the same survey which was conducted in a rural school in County Roscommon?

Q,13 An electronic manufacturer wants to sell a new MP3 player with one hundred songs preloaded on to it. They decide that by preloading one hundred of the most popular songs they will have a best selling product when it is released onto the market. The company is based in France and they survey thirty thousand French people to determine which one hundred songs should be loaded onto the MP3 player. They then release the player in every country in the world.

Do you think that this MP3 player will be popular in Ireland? In the USA? In France? In Belgium?

Q.14 The European Union (EU) wants to standardise timekeeping and establish one time-zone for the entire EU. Before bringing forward legislation the EU decides it should gauge public opinion by carrying out a survey of EU citizens.

The population of the EU is approximately five hundred million people, so the EU decides that fifty thousand people will be surveyed.

a. How should the EU choose the fifty thousand people to participate in the survey?

b. One person suggests that to save money the survey could easily be conducted by stopping people outside the EU offices in Luxembourg and asking their opinion. Is this an appropriate method?

c. Another person indicates that the sample should be chosen as a simple random sample, so that the EU should place the names and addresses of every person in the EU over the age of 18 into a computer and then randomly select fifty thousand people who would be written to and surveyed. Is this simple random sampling approach one that you would recommend?

d. How would the principle of stratified random sampling be applied to this survey? Would it be preferable to simple random sampling?

Q.15 A company owns fifty shops, each of which is the same size and shape and stocks the same selection of products. Each shop contains ten aisles and each aisle contains different brands of one single product. For example, aisle number 1 in each shop contains 40 inch LCD televisions and aisle 2 contains packets of tea.

The company manager wants to conduct a survey of the stores to determine which brands sell best but he does not want to survey each item in every store. Instead, he wants to select a sample of items. The manager has heard that cluster sampling and stratified sampling are good techniques to use but he does not understand the difference between the two methods.

Explain how each method could be applied to select a sample of products from the company's shops. Discuss the relative merits of each sampling method in this context.

Q.16 You are asked to conduct a survey of students in your school to decide if everyone should wear a new school uniform that has a red and black striped design. You have a picture of the new uniform and you must choose a sample of fifty students to show the picture to. Design four different plans to select the students for your sample, one for each of the following different sampling strategies:

(a) Cluster Sampling, (b) Simple Random Sampling,
(c) Stratified Random Sampling, (d) Quota Sampling.

Q.17 Construct a sampling plan to decide what proportion of sweets in packets of M and Ms are yellow. Indicate why the sampling scheme you have chosen is the best possible scheme and will yield the most accurate prediction.

Q.18 Suppose that RTE wants to know how viewers feel about a new soap opera they are broadcasting. They decide that after the show they will show two phone numbers for people to call one for people who like the show and one for people who don't like the show. Will this result be biased? Explain in one sentence.

Q.19 Suppose your school principal wants to know how students feel about a policy on banning students being dropped by car to school. The principal wants to make all students travel by foot, on bicycle or by public transport. The principal doesn't want to ask all students so s/he must rely on a sample. Which of the following sampling methods would you recommend that the principal choose?

a. Walk through the school building and pick a classroom at random, choose the students inside as the sample.

b. Place a list of all students in the school on a table and randomly choose 40 students.

c. Stand outside the school in the morning time and stop every 5th student who arrives by car.

d. Randomly pick 5 students from each class in the school.

e. Stop one school bus outside the school in the morning and question all the students on board the bus prior to their disembarkation.

Did you recommend more than one method? Why did you prefer the methods that you chose? What quality were you looking for in the sample you chose? Could you recommend a better sampling scheme than any of the methods listed above?

Q.20 In order to survey the opinions of its passengers an airline made a list of all its flights and randomly selected 25 flights. All the passengers on those flights were asked to fill out a survey. What kind of sampling procedure was used in this case?

Q.21 Pre-election polls are often conducted by asking the opinions of a few thousand adults nationwide and using the results to infer the opinions of all the adults in the nation. Explain what the sample is and what the population for such polls is.

Q.22 In order to find out how people connected to your school feel about the re-introduction of undergraduate fees for third level education. Your principal considered three different categories of people: teachers in your school, students in your school and parents of students in your school. A random sample from each group was surveyed. What kind of sampling procedure was used in this case?

Q.23 An experiment is conducted to measure the effectiveness of step aerobics as a means of weight loss. What would be the explanatory variable and the outcome variable in this experiment?

List two reasons why it might be preferable to conduct a sample survey rather than a census.

Q.24 In a designed experiment the sample chosen for the study is often split into a 'treatment group' and a 'control group'. What differences or similarities exist between these two groups?

In a case-control study the sample chosen for the study is split into a 'case group' and a 'control group'. How do these two groups differ from the 'treatment group' and 'control group' used in a designed experiment?

A study wishes to examine whether there is a link between baldness and susceptibility to having a heart attack for males. Recommend how a case control study could be used to answer this question. Why is it more appropriate to use 'case and control' rather than 'treatment and control' as the categories for dividing the sample group in this study?

STATISTICS 4

SYLLABUS TOPIC: ANALYSING, INTERPRETING AND DRAWING INFERENCES FROM DATA

LEARNING OUTCOME

As a result of completing the activities in this section you will be able to

- recognise how sampling variability influences the use of sample information to make statements about the population.
- develop appropriate tools to describe variability drawing inferences about the population from the sample.
- interpret the analysis
- relate the interpretation to the original question

Students working at OL will be able to

- discuss different types of studies: sample surveys, observational studies and designed experiments
- select a sample (Simple Random Sample)
- recognise the importance of representativeness so as to avoid biased samples
- design a plan and collect data on basis of above knowledge

INTRODUCTION

Throughout the activities you will analyse and interpret the data from a statistical investigation and draw inferences and conclusions based on your analysis of the data.

Q.1 In a school there are two 6th year Mathematics teachers called Holly and Luke. Over the last ten years the mean mark achieved by Holly's students on the Leaving Certificate Mathematics Exam was 60%, with a standard deviation of 2%. During the same period Luke's students also have a mean mark of 60% but the standard deviation of these students' marks is 15%.

Students in this school have a choice of teacher when entering 6th year. Which teacher should a student who is good at mathematics choose? Which teacher should a student who is weak at mathematics choose?

Q.2 Indoor radon concentrations in Ireland, X, are measured and it is found that the logs of these concentrations, Y=log(X), follow an approximate normal distribution. When 20,000 radon values were considered it was found that the mean indoor radon concentration was 89 Becquerels per cubic metre (Bm^{-3}). Using the information provided in this question would 200 Bm^{-3} or 20 Bm^{-3} be a more likely value for the standard deviation of these radon concentrations?

Suppose that a house is found with an indoor radon concentration of 40,000 Bm^{-3}. Taking the mean and standard deviation above, and taking into account the normality of the Y values, do you think that there is anything unusual about this house? Would you consider a measurement of 8,000 Bm^{-3} to be an outlier for this data set?

Q.3 Two hundred athletes enter a race. One athlete wins the race with a time that is very much faster than any of the other athletes. The athlete is accused of using performance enhancing drugs. At a subsequent court hearing the judge becomes frustrated with the cases put forward by the defence and the prosecution. The judge calls you as an expert witness to use your knowledge of statistics to try to establish the truth. What do you do? How would you analyse the data to establish the innocence or guilt of the accused athlete?

Q.4 One thousand people are selected using simple random sampling to take part in a survey. The participants are blindfolded and are asked to taste two different brands of crisps and to indicate which of the two brands, A or B, they prefer. What is the margin of error for this survey? If 60% of people in the sample indicated a preference for Brand A, what can you say about the proportion of individuals in the population that preferred brand B?

How many individuals should participate in a survey to achieve a margin of error of 2%?

Q.5 The Blue Party have been the party in power for the last 4 years. The leader of the Blue Party, knows that he will have to face the electorate at some time during the next year. He can call an election today or wait for up to 12 months before he must call an election. He wants to maximise the chances that his Blue Party is re-elected to government, so he asks a polling company to conduct an opinion poll of the electorate. The company is instructed not to allow the knowledge of this poll to become known so it decides to survey just 100 people rather than the usual 1000. The survey indicates that 56% of those polled say they will vote for the Blue Party. The sample was chosen using simple random sampling to be representative of the entire electorate. Should the leader of the Blue Party call an election on the basis of the survey results?

Q.6 A national newspaper publishes the results of an opinion poll on the 22nd of May indicating that 45% would vote for the Democratic People's Party. The following week a different national newspaper reports that a new poll shows 42% of people would now vote for the Democratic People's Party. This newspaper prints the headline 'Significant decline in support for Democratic People's Party'. If a simple random sample of 1000 people were surveyed in each of the two polls, what is your opinion of the newspaper's headline?

Activity 4.1

Do taller people have longer arms than shorter people?

Select 20 students from your school and line them up from the shortest to the tallest. Now hand a piece of paper with the letter A to each of the 10 shortest people and a B to the 10 tallest people. Now ask everyone to hold their arms out horizontally and measure the length of each person's arms from the tip of the middle finger on their left hand to the tip of the middle finger on their right hand, measuring across their backs.

Write the length that you have measured on the piece of paper you previously gave to each person. Gather up each of the pieces of paper and order them from the one with the smallest 'arm length' to the one with the longest 'arm length'. Now starting with the first piece of paper write down whether it contained an A or a B, continue writing down As or Bs for each piece of paper until you have a sequence of As and Bs.

Now count the number of As before the first B and add it to the number of Bs after the last A to get a value for the 'Tukey Test Statistic'.

Now making use of the Tukey Quick Test Tables answer the original question: 'Do taller people have longer arms than shorter people?'

Q.7 A company believes it has invented a new drink which will make people better able to understand statistics. The company recruits twenty students studying for a Statistics degree to take part in this experiment. The company does not see the point in using control groups and treatment groups and so it gives its new drink to all the students and measures the students' ability to answer statistical questions. The students all perform well so the company believes it has proven that its drink can improve peoples' statistical understanding.

Comment on the validity of this study.

How would you design a different study to test the effectiveness of the new drink making use of the concept of control groups and treatment groups?
Would this study have any benefits over the study that was conducted by the company? If so, what are the benefits?

Q.8 It is believed that randomisation plays an important role in designing studies to test the effectiveness of a new product. Why is randomisation important?

Consider the case of a company which has developed a shampoo that it claims will make men irresistible to women. To test this claim the company's female scientists recruit forty male test subjects. The female scientists choose some of the male subjects because they consider the males to be particularly handsome, other males are chosen by the female scientists because they consider these males to be particularly unappealing.

The test will involve assigning twenty male test subjects to a treatment group who will wash their hair with the new shampoo every day for a week. The other twenty males will wash their hair with an ordinary shampoo chosen by the scientists. At the end of the week both groups will be sent to a reality television show where a judging panel of five celebrity females will rate each male as to their desirability.

a. How should the forty males be assigned to the treatment and control groups to test the effectiveness of this new product?
b. Why would randomisation play an important role in this assignment?
c. If randomisation were not employed in this study what problems could arise in determining the effectiveness of the shampoo?
d. Suppose that twenty of the males were initially classified as handsome and twenty were classified as unappealing by the female scientists. One of the scientists recommends that in this situation it might be possible to assign the males in a non-random manner to the control and treatment groups. The scientist says that in this particular situation she can produce a study design which is better than the study design that uses randomisation. Is she right? Why? What do you believe is her improved non-random assignment method?

Q.9 Do men or women have more friends?

To answer this question a researcher asked 10 men and 10 women to list the names of their friends whom they had spent time with during the past year. The researcher then counted the numbers of friends for each woman and each man. The results are given below:

Men: 5, 9, 10, 4, 13, 22, 21, 19, 28, 17
Women: 35, 21, 10, 18, 9, 18, 23, 42, 33, 29

Apply the Tukey Quick Test to answer the original question: 'Do men or women have more friends?'

Q.10 A student was given two lists of numbers by his teacher and asked to perform the Tukey Quick Test to decide which of the lists contains the larger numbers.
The two lists of numbers were:

List A: 10, 12, 40, 56, 23, 34, 43, 19, 21,
List B: 8, 10, 34, 36, 23, 36, 39, 17, 9.

The student refused to perform the Tukey Quick Test, saying instead that it was obvious that the numbers in list A were larger than the numbers in List B and that there was no point in performing the Tukey Quick Test.

The teacher was dismayed at the student's refusal and reported him to the principal, where the student still steadfastly refused to conduct the Tukey test, insisting that it was obvious that the numbers in List A were larger than those in List B. The student was expelled from school for subordination. What do you think of the student's position? If you were a statistically literate judge at the subsequent court case, whose side would you rule in favour of, the school or the student? Why?

Q.11 What is the point of performing a Hypothesis test? In making decisions based on samples of data, why can't we just calculate means or medians for each sample and compare the means or medians to reach a conclusion?

STATISTICS 5

SYLLABUS TOPIC: REPRESENTING DATA GRAPHICALLY AND NUMERICALLY

LEARNING OUTCOME

As a result of completing the activities in this section you will be able to

- use scatterplots to determine the relationship between variables (OL)
- recognise that correlation is a value from -1 to +1 and that it measures the extent of linear relationship between two variables (OL)
- match correlation coefficient values to appropriate scatter plots (OL)
- draw the line of best fit by eye (HL)
- make predictions based on the line of best fit (HL)

Activity 5.1

Consider the following:

The Proprietor of Mike's Machines Garage was anxious to encourage his customers to regularly change the oil in their vehicles. He surveyed his customers over a period of time and recorded the data in the table below.

Oil Changes per Year	3	5	2	3	1	4	6	4	3	2	0	10	7
Cost of Repairs (€)	300	300	500	400	700	250	100	400	450	650	600	0	150

a. Represent this data in a graph.

Think: How should the axes be labeled? Should you include units on the axes? What scale should you use on each axis?

b. Visualise a straight line as a representation of the data. Draw a line that seems to 'fit' the plotted points. [A line that 'fits' the points should have the same characteristics as the set of points; it should actually summarise the data. A line drawn this way is called a line of best fit by eye.]

c. Look at the points you plotted and the line of best fit. Do you think there is a relationship?

If so:

- Is it a positive or negative relationship? (Look at how the line 'slopes' to help you decide.)
- Is it a strong or weak relationship? (Look at how close the points are to the line of best fit to help you decide.)

Is there a correlation between the cost of repairs and the number of oil changes per year? If so, describe this correlation.
Do you think the number of oil changes is the only factor which causes the need for engine repairs?
Think of other factors that might affect the need for engine repairs.

It is important to understand that a relationship doesn't mean a cause.

d. Would this data convince Mike's customers that they should change their oil regularly? Give a reason for your answer.

e. According to Mike's data what is the optimum number of oil changes per year?

f. By choosing some points on the line you have drawn that best fits the data, calculate the equation of this 'line of best fit'.

g. Use your graph to predict how much a customer is likely to spend on repairs if they change the oil 5 times a year.

h. Comment on your answer to g) above; is it good value to change the oil 5 times in the year.

i. Examine the slope of the graph; you can do this by counting units.

It is an important learning outcome that you can interpret the slope as 'rate of change'.
What does a 'unit' represent for the number of oil changes per year?
What does a 'unit' represent for the cost of engine repairs?

Complete the sentence below, describing the change in the cost of repairs with the number of oil changes.

A rate of change (or slope) indicates that, for each additional oil change per year, the cost of engine repairs will tend to

Q.1 The chart below is used by veterinary surgeons to decide the dose of a certain drug to be used in fighting bacterial infection in the joints of small animals.

Animal's Weight (kg)	Usual dosage (mg)	Max dosage (mg)
0.9	9	13.5
1.4	14	21
1.8	18	27
2.3	23	34.5
2.7	27	40.5
3.2	32	48
3.6	36	54
4.1	41	61.5
4.5	45	67.5
5.0	50	75
5.4	54	81
5.9	59	88.5
6.4	64	96
7.7	77	115

1. Use graph paper to plot the data (animal weight, usual dosage) and draw a line of best fit.
2. Plot (animal weight, maximum dosage) on the same axes. Draw a line of best fit.
3. Find the slope for each line. What do they mean, and how do they compare?
4. By choosing two points on your lines of best fit write an equation for each of the two lines.
5. Are the two lines parallel? Why or why not?

Note: The data is real data obtained from the drug company website.

Q.2 In BMX dirt-bike racing, jumping high or 'getting air' depends on many factors: the rider's skill, the angle of the jump, and the weight of the bike. Here are data about the maximum height for various bike weights.

Weight (kg)	Height (cm)
8.6	26.3
8.8	26.2
9.1	26
9.3	25.9
9.5	25.7
9.8	25
10.4	24.9
10.2	24.9
10.7	24.6
10.9	24.4

1. Use graph paper to plot the data (weight, height). If the data show a linear relationship, draw a line of best fit.
2. Explain the relationship between bike weight and jump height by completing the following sentence:

   As the weight of the bike increases

3. Use two points on the line of best fit to find the slope or rate of change. What does this mean?
4. Predict the maximum height for a bike that weighs 9.8kg if all other factors are held constant.

Note: Data obtained from BMX magazine article.

Q.3 Use the data provided to help you decide whether or not there is a relationship between the Total Fat content of fast food and the number of Total Calories in the food.

Sandwich	Total Fat (g)	Total Calories
Hamburger	9	260
Cheeseburger	13	320
Quarter Pounder	21	420
Quarter Pounder with Cheese	30	530
Big Mac	31	560
Sandwich Special	31	550
Sandwich Special with Bacon	34	590
Crispy Chicken	25	500
Fish Fillet	28	560
Grilled Chicken	20	440
Grilled Chicken Light	5	300

Q.4 A student was investigating whether students who study more watch less television.

The results obtained are displayed below.

Number of hours per week spent watching television	Number of hours per week spent studying
33	10
28	10
27	12
22	11
28	13
25	15
21	17
23	20
19	20
18	20
21	25
17	25
12	25
7	28
4	30
11	30
22	30
10	32
8	32

Draw a suitable graph and use it to decide whether or not there is a correlation between the time per week spent studying and the time per week spent watching television. Describe the correlation, if there is one.

The students who gathered this information said there is a correlation between the two variables. They said the equation of the line of best fit is; $y = x + 40$.
Look at your graph and explain how you can tell that this equation is wrong.
Draw a line of best fit.
Choose any two points on this line and use them to find the equation of the line of best fit.

A student who was absent on the day of the investigation reported that he spends 6 hours per week watching TV. Use the information the students gathered to predict how many hours that student spends studying /week.

Is it possible that your predicted answer is incorrect? Explain your thinking.

Q.5 (a) Which implies a stronger linear relationship, a correlation coefficient of +0.2 or –0.5? (b) Is the correlation between the following pairs of variables likely to be strong, moderate, or weak? Is it likely to be positive or negative?

a. Daily rainfall in Sydney and Dublin
b. The number of hours that students spend studying during three weeks prior to their Leaving Certificate Mathematics Exam and their percentage score on the exam
c. Daily rainfall in Drogheda and Dundalk
d. Engine size of a car and its petrol consumption
e. Weight of a randomly selected woman and the amount of food she eats
f. The average daily temperature for different towns in Europe and the average cost of heating homes in the same towns
g. Number of students in University College Dublin who drive cars to college each day and number of free car parking spaces in the university each day

Q.6 Is there a relationship between a student's height and their shoe size? Select twenty students from your school and make a note of the height (X) and shoe size (Y) for each student.

Draw a scatterplot of the data. Does there seem to be a relationship between the two variables?
Compute the correlation coefficient for this set of data and use this value to answer the question as to whether there is a relationship between a student's height and their shoe size.

When you select students to participate in this study should you choose students who are all of similar heights? If your school had both girls and boys would it make sense to have both in your study or should you limit your group to just males or females?

Q.7 The shelf life of packaged food depends on many factors. No one likes soggy cereal so it is clear that moisture content is important in determining the shelf life of cereal. Statistics students with part-time jobs in supermarkets conducted an experiment on one particular brand of cereal. They recorded time on shelf X (days), and moisture content Y (percentage). The table below shows the data they collected.

X	0	3	6	8	10	13	16	20	24	27	30	34	37	41
Y	4.1	4.3	4.4	4.9	2.8	3.0	3.1	3.2	3.4	3.4	3.5	3.1	3.8	4.0

i. Construct a scatterplot for this data. From the scatterplot does it seem that there is a relationship between the length of time that a package spends on the shelf and the moisture content of the cereal package?

ii. Compute a correlation coefficient for this data set and interpret the value of the correlation coefficient in the context of this data.

Q.8

a. University researchers conducted a study looking at children under the age of 12 and they found that there was a strong positive correlation between the numbers of fillings in children's teeth and the children's vocabulary. Does this mean that eating more sweets would increase a child's vocabulary? Explain.

b. These researchers also conducted a study where they examined each country in the world and they found that there was a strong positive correlation between the number of storks in a country and the number of babies born in that country. When a newspaper discovered this information they had a front page headline which read 'Researchers have shown that storks really are responsible for bringing babies'. Explain the error made by the journalist. Can you identify a possible explanation for the researchers' result?

c. The same researchers also found a strong positive correlation between the sales of ice-cream in Ireland and the number of people who drowned in Ireland for each week of the year. Does this mean that consuming ice-cream increases the likelihood that someone will drown? Suggest an explanation for the result the researchers found.

Q.9 The following information was obtained from the manager of a local Water Department for predicting the weekly consumption of water in litres from the size of household:

Household Size	2	7	9	4	12	6	9	3	3	2
Water Used	650	1200	1300	430	1400	900	1800	640	793	925

a. Without performing any computations, predict what the correlation coefficient would be for this set of data.

b. Which variable should be labelled X and which should be labelled Y in a scatter plot of this data?

c. Construct a scatter plot of this data and determine whether there is a relationship between household size and water consumption.

d. By looking at the scatter plot decide whether you believe your previous estimate of the correlation coefficient. If necessary make a new prediction based on the scatter plot.

e. Compute the correlation coefficient for this data set. Does the computed correlation coefficient match with your previous two predictions?

f. Interpret the value of the correlation coefficient that you computed. Suppose that, instead of measuring the water consumption in litres, the engineers had measured the water consumption in gallons. If you were to convert the values in the table above from litres to gallons what effect would this have on the correlation coefficient?

Q.10 Which of the following is not a property of correlation?

a. A negative correlation indicates that the variables increase together.

b. A correlation will be between -1 and $+1$.

c. Correlations are not affected by changes in units of measurement.

d. A correlation of zero indicates that there is no linear relationship between the two variables