Genetics and Mathematics

Key Stage: 4

Strand: Data Handling

Learning Unit: More about probability

Objective: Understand how to apply mathematics to verify Hardy Weinberg Principle in genetics

Prerequisite Knowledge:

(i) Recognise the concept of probability and be able to calculate probabilities of events by listing the sample space and counting

(ii) Use tables or tree diagrams to list the sample space

(iii) Understand the multiplication law of probability and the concept of independent events

(iv) Perform operations of algebraic fractions

Relationship with other KLA(s) in STEM Education:

Compulsory Part II. Genetics and Evolution – Basic genetics in *Science Education Key Learning Area Curriculum and Assessment Guide: Biology (Secondary 4 - 6) (with updates in November 2015)*

Description of the Activity:

1. To start with the common generic traits in human
2. To give examples of generic traits that are either dominant or recessive
3. To recognise inheritance patterns and Punnett Squares
4. To verify Hardy-Weinberg Principle, which states that genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences

Activity 1: Traits of our Class

Some traits are common in a population, but an individual's overall combination of traits makes us unique. Two students form a group. Working in pairs, students observe each other and take an inventory of their own traits. They record their observations in a table and find out the most and least common traits in the class.

How many traits do you have? Complete the following table to find out. Put a tick (✓) in the appropriate box.

	Trait	Yes	No
(i)	Can roll his or her tongue		
(ii)	Thumb bending		
(iii)	Double eyelid		
(v)	Index finger is longer than ring finger		
(vi)	Has dimples		
(vii)	Has detached earlobes		
(xi)	Naturally curly hairs		
(xii)	Is left-handed		
(xiii)	The left thumb is on top (L) in Hand Clasping		

The teacher records the data by counting the total number of students who marked "Yes" and the total number of students who marked "No" for each trait. Students can calculate the frequency of each trait in the whole class, and then compare their calculated frequencies with those <u>for the general population</u> (see Table 1).

Trait	**Frequencies**		
Tongue rolling	Can roll tongue – 70%	Cannot roll tongue – 30%	
Handedness	Right handed – 93%	Left handed – 7%	
Hand clasping	Left thumb on top – 55%	Right thumb on top – 44%	No preference – 1%

Table 1: Frequencies of traits in the general population*

***Frequencies for traits are from Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim/)**

<u>Activity 2: Recognise inheritance patterns</u>

Background information:

Traits in children are determined by inherited genes from parents (e.g. dimples and double eyelid). Suppose the inherited traits are controlled by a pair of genes only. Each gene comes in just two different versions, namely **dominant** (denoted by *A*) and **recessive** (denoted by *a*). When dominant gene and recessive gene co-exist, only the trait due to dominant gene will be shown out.

There are two kinds of genes.

Dominant gene (Denoted by *A*)	When dominant gene and recessive gene co-exist, the characteristics described by the dominant gene will be shown out. For example, *Aa* will show out the characteristics described by the dominant gene.
Recessive gene (Denoted by *a*)	When dominant gene and recessive gene co-exist, the characteristics described by the recessive gene will NOT be shown out. Therefore, only *aa* will show out the characteristics described by the recessive gene.

The combination of genes is called the "genotype". There are 3 kinds of genotypes.

AA	Both are dominant genes
Aa	One dominant gene and one recessive gene
aa	Both are recessive genes

Examples of human single-pair genetic inheritance:

Traits of dominant genes	**Traits of recessive genes**
Curly tongue	Non-curly tongue
Separated earlobe	Attached earlobe
Index finger is longer than ring finger	Ring finger is longer than index finger
Dimple	No dimple
Straight thumb	Thumb bending
Double eyelid	Single eyelid

Given parental genotypes, we can construct the possible combinations of genotypes of the child. The first two have been done for you as an example.

F \ M	*A*	*a*
A	*AA*	*Aa*
a	*Aa*	*aa*

F \ M	*A*	*A*
A	*AA*	*AA*
a	*Aa*	*Aa*

F \ M		

F \ M		

F \ M		

Note: M – Mother; F – Father; *A* – Dominant genes; *a* – Recessive genes

There are three kinds of parental genotypes: *AA*, *Aa* and *aa*. The number of combinations is 6 instead of $3 \times 3 = 9$ as the potential genotypes for some combinations of the child are the same. For example, the potential genotypes of the child under the combination of father genotype *AA* with mother genotype *Aa* is the same as that under the combination of father genotype *Aa* with mother genotype *AA*.

Based on the result in the completed Punnett squares on page 4 and suppose that each gene in father combines randomly with each gene in mother, please find out the probabilities of child's genotypes. The first two have been done for you as an example.

Parental genotypes	*AA* and *AA*	*AA* and *Aa*	*AA* and *aa*
Probability of child's genotype for each combination	$P(AA)=1$ (All are *AA*)	$P(AA)=\frac{2}{4}=\frac{1}{2}$	$P(AA)=$
	$P(Aa)=0$	$P(Aa)=\frac{2}{4}=\frac{1}{2}$	$P(Aa)=$
	$P(aa)=0$	$P(aa)=\frac{0}{4}=0$	$P(aa)=$

Parental genotypes	*Aa* and *Aa*	*Aa* and *aa*	*aa* and *aa*
Probability of child's genotype for each combination	$P(AA)=$	$P(AA)=$	$P(AA)=$
	$P(Aa)=$	$P(Aa)=$	$P(Aa)=$
	$P(aa)=$	$P(aa)=$	$P(aa)=$

Questions:

1. If the father has two recessive genes (*aa*) and the mother has one dominant gene and one recessive gene (*Aa*), what is the probability of child's genotype with one dominant gene (*AA* or *Aa*)?
2. If the father has one dominant gene and one recessive gene (*Aa*) and the mother also has one dominant gene and one recessive gene (*Aa*), what is the probability of child's genotype with two recessive genes (*aa*)?

Suggested Solutions:

1. $\frac{2}{4}=\frac{1}{2}=50\%$
2. $\frac{1}{4}=25\%$

Task 3: Proof of Hardy-Weinberg principle

Background:

Will the population with characteristics represented by the dominant gene increase after several generations, and the population with characteristics represented by the recessive gene disappear after a hundred years? G.H. Hardy, a British Mathematician in year 1908, showed that this phenomenon would NOT occur, by using the knowledge of probability and algebra in Mathematics. Hardy with Weinberg, who was a biologist, published this result in a magazine and it is called the **Hardy-Weinberg principle.**

The teacher may do the formal proof with the students in the following way.

Assume that the probabilities of the combinations of parental genotypes are P(*AA*) = p_1, P(*Aa*) = p_2 and P(*aa*) = p_3, where $p_1 + p_2 + p_3 = 1$.

Parental genotypes	Probability of this combination of parental genotypes	Probability of child's genotype		
		AA	*Aa*	*aa*
AA and *AA*	$(p_1)^2$	1	0	0
AA and *Aa*	$p_1p_2 + p_2p_1 = 2p_1p_2$	$\frac{1}{2}$	$\frac{1}{2}$	0
AA and *aa*	$p_1p_3 + p_3p_1 = 2p_1p_3$	0	1	0
Aa and *Aa*	$(p_2)^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Aa and *aa*	$p_2p_3 + p_3p_2 = 2p_2p_3$	0	$\frac{1}{2}$	$\frac{1}{2}$
aa and *aa*	$(p_3)^2$	0	0	1

According to the above table, the probabilities of child's genotypes are as follows.

P(*AA* in child's genotype) = $p_1^2 \times 1 + 2p_1p_2 \times \frac{1}{2} + p_2^2 \times \frac{1}{4} = p_1^2 + p_1p_2 + \left(\frac{1}{2}p_2\right)^2 = \left(\frac{2p_1 + p_2}{2}\right)^2$

P(*Aa* in child's genotype) = $2p_1p_2 \times \frac{1}{2} + 2p_1p_3 \times 1 + p_2^2 \times \frac{1}{2} + 2p_2p_3 \times \frac{1}{2} = \frac{(2p_3 + p_2)(2p_1 + p_2)}{2}$

P(*aa* in child's genotype) = $p_2^2 \times \frac{1}{4} + 2p_2p_3 \times \frac{1}{2} + p_3^2 \times 1 = \left(\frac{2p_3 + p_2}{2}\right)^2$

In what condition will give rise to a stable genotype distribution?

A stable genotype distribution means that: the ratio of the population with characteristics represented by the dominant gene, to the population with characteristics represented by the recessive gene, will become constant. That is to say, **the child's genotype distribution will be the same as the parental genotype distribution**, or the probabilities of the three genotypes of the child is the same as those of the parental genotypes:

$$p_1 = \left(\frac{2p_1 + p_2}{2}\right)^2 \text{ and } p_3 = \left(\frac{2p_3 + p_2}{2}\right)^2 \quad \ldots (*)$$

Since $p_2 = 1 - p_1 - p_3$, (*) already guarantee the child's probability of *Aa* genotype is the same as that of the parent generation.

Suppose (*) is true, we have

$$\begin{aligned}\sqrt{p_1} + \sqrt{p_3} &= \frac{2p_1 + p_2}{2} + \frac{2p_3 + p_2}{2} \\ &= \frac{2p_1 + p_2 + 2p_3 + p_2}{2} \\ &= \frac{2(p_1 + p_2 + p_3)}{2} \\ &= 1\end{aligned}$$

On the other hand, if $\sqrt{p_1} + \sqrt{p_3} = 1$, can we ensure that (*) is true, i.e., will **the child's genotype distribution be the same as the parental genotype distribution**?

Based on students' abilities, the teacher may provide hints about squaring both sides of $\sqrt{p_1} + \sqrt{p_3} = 1$ and use the relation $p_1 + p_2 + p_3 = 1$ to carry out the investigation.

Actually,

$$\left(\sqrt{p_1} + \sqrt{p_3}\right)^2 = 1^2$$

$$p_1 + 2\sqrt{p_1 p_3} + p_3 = 1$$

$$2\sqrt{p_1 p_3} = 1 - p_1 - p_3$$

$$2\sqrt{p_1 p_3} = p_2$$

$$4p_1 p_3 = p_2^2$$

$$p_1(1 - p_1 - p_2) = \frac{p_2^2}{4}$$

$$p_1 = p_1^2 + p_1 p_2 + \frac{p_2^2}{4}$$

$$p_1 = \left(p_1 + \frac{p_2}{2}\right)^2$$

$\therefore \quad p_1 = \left(\frac{2p_1 + p_2}{2}\right)^2$. Similarly, we can derive another relation $p_3 = \left(\frac{2p_3 + p_2}{2}\right)^2$.

Hence, if $\sqrt{p_1}+\sqrt{p_3}=1$, (*) is true. That is to say, if a generation with p_1 and p_3 satisfying the relation $\sqrt{p_1}+\sqrt{p_3}=1$, the genotype distribution in the next generation will the same as the genotype distribution of the original generation.

Interestingly, we observe that no matter p_1 and p_3 satisfy $\sqrt{p_1}+\sqrt{p_3}=1$ or not, we have

$$\sqrt{P(AA\text{ in child's genotype})}+\sqrt{P(aa\text{ in child's genotype})}$$

$$=\sqrt{\left(\frac{2p_1+p_2}{2}\right)^2}+\sqrt{\left(\frac{2p_3+p_2}{2}\right)^2}$$

$$=\frac{2p_1+p_2}{2}+\frac{2p_3+p_2}{2}=p_1+p_2+p_3=1$$

$$\therefore \sqrt{P(AA\text{ in child's genotype})}+\sqrt{P(aa\text{ in child's genotype})}=1.$$

This result is still true for different values of p_1, p_2 and p_3.

Given the first generation with p_1 and p_3. Even though $\sqrt{p_1}+\sqrt{p_3}\neq 1$, we still have $\sqrt{P(AA\text{ in child's genotype})}+\sqrt{P(aa\text{ in child's genotype})}=1$ for the second generation. Afterwards, the relation $\sqrt{P(AA)}+\sqrt{P(aa)}=1$ holds for each subsequent generation. As a result, the genotype distribution in the third generation is the same as that in the second generation. The genotype distribution in the fourth generation is the same as that in the third generation, and so on. The next generation of the child's genotype is thus **a stable genotype distribution**.

This relation states that whatever the value of p_1 and p_3 are at the beginning, the child's genotype distribution in the future generations becomes **stable**. In other words, the population with traits represented by the recessive gene **will not** disappear after hundred years.

Some students are more familiar with numerical computation. The teacher may guide the students to complete the following three practices before attempting the proof of Hardy-Weinberg principle. The teaching efficiency may thus be enhanced.

Practice 1

Assume that the probabilities of the combinations of parental genotypes are P(*AA*) $= \frac{1}{2}$ and P(*Aa*) $= \frac{1}{3}$.

(a) What is the value of P(*aa*)?

P(*aa*) = ______________________________

Once the values of P(*AA*) and P(*Aa*) are known, we can determine the value of P(*aa*) by $1 - P(AA) - P(Aa)$.

(b) Complete the following table.

Parental genotypes	Probability of this combination of parental genotypes	Probability of child's genotype		
		AA	*Aa*	*aa*
AA and *AA*		1	0	0
AA and *Aa*		$\frac{1}{2}$	$\frac{1}{2}$	0
AA and *aa*				
Aa and *Aa*				
Aa and *aa*				
aa and *aa*				

(c) Using the above table, calculate the following probabilities:

(i) P(*AA* in child's genotype)

= P("parental genotype is *AA* and *AA*") × P("*AA* in child's genotype" | "parental genotype is *AA* and *AA*")

+ P("parental genotype is *AA* and *Aa*") × P("*AA* in child's genotype" | "parental genotype is *AA* and *Aa*")

+ P("parental genotype is *Aa* and *Aa*") × P("*AA* in child's genotype" | "parental genotype is *Aa* and *Aa*")

$= \frac{1}{4} \times 1 + \frac{1}{3} \times \frac{1}{2} + \frac{1}{9} \times \frac{1}{4}$

$= \frac{4}{9}$

(ii) P(*Aa* in child's genotype)

=

(iii) P(*aa* in child's genotype)

=

Practice 2

Assume that the probabilities of the combinations of parental genotype is P(*AA*) = 0.25 and P(*Aa*) = 0.25.

(a) What is the value of P(*aa*)?

P(*aa*) = ____________________

(b) Complete the following table:

Parental genotypes	Probability of this combination of parental genotypes	Probability of child's genotype		
		AA	*Aa*	*aa*
AA and *AA*				
AA and *Aa*				
AA and *aa*				
Aa and *Aa*				
Aa and *aa*				
aa and *aa*				

(c) Using the above table, calculate the following probabilities:

(i) P(*AA* in child's genotype)

(ii) P(*Aa* in child's genotype)

(iii) P(*aa* in child's genotype)

Practice 3

We will do two more sets of calculations by using a spreadsheet software such as Microsoft Excel.

(a) Assume that the probabilities of the combinations of parental genotypes are P(*AA*) = 0.1 and P(*Aa*) = 0.7. Write down the following values:

P(*AA* in child's genotype) = ____________________

P(*Aa* in child's genotype) = ____________________

P(*aa* in child's genotype) = ____________________

What is the value of $\sqrt{P(AA\text{ in child's genotype})}+\sqrt{P(aa\text{ in child's genotype})}$? __________

(b) Assume that the probabilities of the combinations of parental genotypes are P(*AA*) = 0.9 and P(*Aa*) = 0.09. Write down the following values:

P(*AA* in child's genotype) = ____________________

P(*Aa* in child's genotype) = ____________________

P(*aa* in child's genotype) = ____________________

What is the value of $\sqrt{P(AA\text{ in child's genotype})}+\sqrt{P(aa\text{ in child's genotype})}$? __________

(c) In practice 2, what is the value of $\sqrt{P(AA\text{ in child's genotype})}+\sqrt{P(aa\text{ in child's genotype})}$?

____

In practice 3, what is the value of $\sqrt{P(AA\text{ in child's genotype})}+\sqrt{P(aa\text{ in child's genotype})}$?

____

References:

1. 蕭文強、林建（2010）。*概率萬花筒（數學百子櫃系列（八））*。香港：教育局課程發展處數學教育組。
2. 鄭惟厚（2007）。*你不能不懂的統計常識*。台北：天下文化。
3. https://teach.genetics.utah.edu/content/heredity/files/InventoryOfTraits.pdf