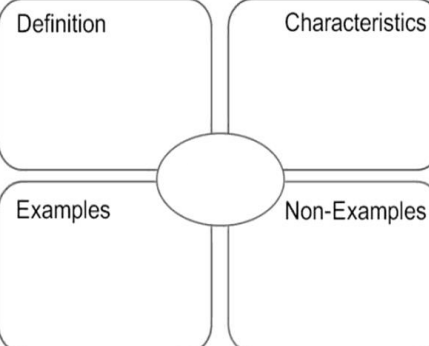


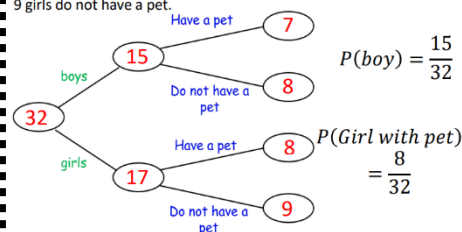
STATS

Summer Term

Probability



In Hannah's class there are 32 students.
15 of these students are boys.
7 of the boys have a pet.
9 girls do not have a pet.



When two events, A and B, are **independent**:

$$P(A \text{ and } B) = P(A) \times P(B)$$

When two events, A and B, are **mutually exclusive**:

$$P(A \text{ or } B) = P(A) + P(B)$$

Topics

- Probability scale
- Probability of events
- Tree diagrams
- Venn diagrams

Career Links

Being able to confidently work with data is a great skill to have and has lots of links with a number of careers such as:

- Statistician
- Business Analyst
- Biostatistician
- Healthcare

What do I need to be able to do?

- Calculate simple probabilities
- Predict the number of outcomes
- Draw and interpret probability trees
- Draw and interpret Venn diagrams



1) What is the probability that a bead chosen will be **yellow**. Show the answer on a number line.

$$\text{Probability} = \frac{\text{Number of favourable outcomes}}{\text{Total number of outcomes}}$$

$$P(\text{Yellow}) = \frac{2}{8} = \frac{1}{4}$$



2) How many **yellow** beads would you **expect** if you pulled a bead out and replaced it 40 times?

$$\frac{1}{4} \times 40 = \frac{1}{4} \text{ of } 40 = 10$$

Key Vocabulary

Probability	The chance of something happening
Impossible	Cannot happen
Certain	Will definitely happen
Even chance	Two outcomes have the same chance of happening
Expectation	The amount of times you expect an outcome to happen
Relative frequency	How often something happens divided by all the outcomes
Venn Diagram	Shows the relationship between two or more finite sets
Mutually exclusive	Cannot happen at the same time
Independent	One event does not affect the probability of another event
Dependent	One outcome affects the other

\in means 'element of a set' (a value in the set)

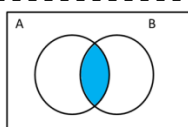
$\{ \}$ means the collection of values in the set.

ξ means the 'universal set' (all the values to consider in the question)

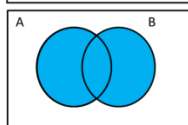
A' means 'not in set A' (called **complement**)

$A \cup B$ means 'A or B or both' (called **Union**)

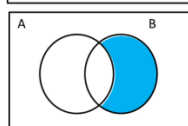
$A \cap B$ means 'A and B (called **Intersection**)



$$P(A \cap B)$$



$$P(A \cup B)$$



$$P(A' \cap B)$$

The probability of an event A happening, **given that** event B has already happened.

With conditional probability, check if the numbers on the second branches of a tree diagram changes. For example, if you have 4 red beads in a bag of 9 beads and pick a red bead on the first pick, then there will be 3 red beads left out of 8 beads on the second pick.



STATISTICS

Summer Term

Statistics

Topics

- Averages from a table
- Stem and leaf
- Scatter graphs
- Sampling

What do I need to be able to do?

- Interpret stem and leaf diagrams
- Draw scatter graphs
- Understand sampling

Career Links

Being able to confidently work with data is a great skill to have and has lots of links with a number of careers such as:

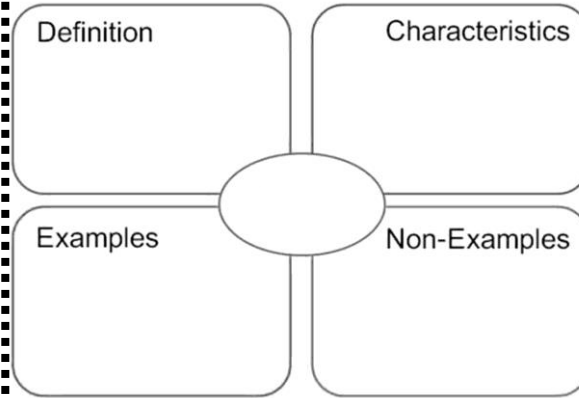
- Statistician
- Business Analyst
- Biostatistician
- Healthcare

Key: 16|2 means 162

Stem	Leaf
15	3 4 4
16	2 5 9
17	2 6 6 6 7
18	0 2 6 7
19	0

Each number is split into two parts.

- The first digit(s) form the stem,
- The last digit forms the leaf.



Average	Advantage	Disadvantage
Mode	Can be used for qualitative data Easy to obtain	There can be more than one mode or even no mode
Median	Not affected by very large or very small values	Can be time consuming when there is a lot of data
Mean	Takes into account all of the data	Very small or very large values affects the mean

Disadvantages of sampling

- Chances of bias.
- Difficulties in selecting truly a representative sample
- Need for subject specific knowledge.
- changeability of sampling units.
- impossibility of sampling.

Averages from Frequency Tables

a) Find the mean of this data

Goals Scored (x)	Frequency (f)	$f \times x$
0	2	$0 \times 2 = 0$
1	2	$1 \times 2 = 2$
2	5	$2 \times 5 = 10$
3	1	$3 \times 1 = 3$
Total	10	15

Step 1: calculate the total frequency
Step 2: calculate $f \times x$
Step 4: calculate the mean

$$\text{Mean} = \frac{\text{Total } fx}{\text{Total } f}$$

$$\frac{\text{Total } fx}{\text{Total } f} = \frac{15}{10} = 1.5 \text{ goals}$$

b) Find the mode

The mode is the one with the highest frequency

Highest frequency = 5

Mode = 2 goals

c) Find the median

$$\text{Median value} = \frac{\text{Total frequency} + 1}{2}$$

$$\frac{11}{2} = 5.5 \text{th value}$$

add the frequency column until you reach the value in-between the 5th and 6th value

Median = 2 goals

d) Find the range

Highest number of goals = 3
Smallest number of goals = 0

$$\text{Range} = 3 - 0 = 3$$

a) Estimate the mean of this data

Length (L cm)	Frequency (f)	Midpoint (x)	$f \times x$
$0 < L \leq 10$	10	5	$10 \times 5 = 50$
$10 < L \leq 20$	15	15	$15 \times 15 = 225$
$20 < L \leq 30$	23	25	$23 \times 25 = 575$
$30 < L \leq 40$	7	35	$7 \times 35 = 245$
Total	55		1095

Step 1: calculate the total frequency
Step 2: find the midpoint of each group

Step 3: calculate $f \times x$

Step 4: calculate the mean

$$\text{Mean} = \frac{\text{Total } fx}{\text{Total } f}$$

$$\frac{\text{Total } fx}{\text{Total } f} = \frac{1095}{55} = 19.9 \text{cm}$$

Scatter Graphs – Explaining Patterns

A weatherman says "Temperatures are higher in towns that have more sunshine". Is this supported by the scatter graph?

= Yes, the majority of points for high temperature appear when there are more hours of sunshine.

Interpolation and Extrapolation

Interpolation – making a prediction of a value that falls within the range of your data. This is more accurate.

Extrapolation – making a prediction of a value that falls outside the range of your data. This is less accurate.

Another town had a maximum temperature of 11°C that day. Use a line of best fit to estimate the hours of sunshine at this town.

Step 1 – Draw a line of best fit = 9.5 hours

Step 2 – Draw a line along from 11°C and down from the line of best fit

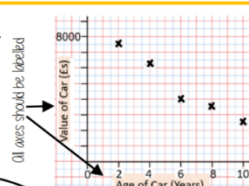
Comment on the reliability of your prediction.

= This is not a reliable estimate because it is extrapolation

Draw and interpret a scatter graph

Age of Car (Years)	2	4	6	8	10
Value of Car (Es)	7500	6250	4000	3500	2500

- This data may not be given in size order
- The data forms information pairs for the scatter graph
- Not all data has a relationship



"This scatter graph shows as the age of a car increases the value decreases"

The link between the data can be explained verbally

The axes should fit all the values on and be equally spread out

Random samples are where each item in the population has an equal chance to be picked. The most common method is assigning each value in the population a number, then randomly picking numbers out of a hat or using a random number generator.

Stratified samples use sub-groups in the population sampled in the same proportion as in the population e.g. If a population of 75 has 20 girls and 40 boys, the sample of 10% (12 students) will have twice as many girls as boys to retain the ratio of girls to boys, therefore 8 girls and 4 boys.

Sampling can be used to approximate the size of a population by doing a capture/recapture method:

e.g. There are an unknown number of birds in a colony. 30 birds are captured and have a tag fitted, then released. The following week a further 30 birds are captured and only 8 have a tag. Approximately how many birds are in the colony?

$$\frac{\text{Sample divided by population (n)}}{\frac{30}{n}} = \frac{8}{30}$$

multiplying by n and by 30 gives us $900 = 8n$

$$\text{Therefore } n = 900 \div 8 = 112.5 = \text{approximately } 113 \text{ birds.}$$