Statistics and Probability

➢ Identifying and Collecting Data

- **Identifying Data**:

  ✓ What is Data?
  Data is information that we collect to help us answer questions or solve problems.

  ✓ Types of Data:
  Data can be numbers (like your age) or words (like your favourite colour). There are two main types: quantitative (numbers) and qualitative (descriptions).

- **Collecting Data**:

  ✓ How to Collect Data?
  There are many ways to collect data. You can use surveys (asking people questions), experiments (testing something and recording what happens), or observations (watching and noting what you see).

  ✓ Why Collect Data?
  We collect data to get information that helps us make decisions, understand things better, and find answers to our questions.

For example, Imagine you want to find out if there is a relationship between the amount of time students spent on homework and their grades.

1. **Identify**: Decide you need data on how much time students spend on homework and their grades.

2. **Collect**:

   - **Design a Survey**: Create a survey with questions like:
     - "How many hours do you spend on homework each week?"
     - "What are your average grades in your subjects?"
   - **Gather Grades**: If possible, collect data on students' grades directly from their report cards (with permission).

3. **Organize the Data**:

   - **Create a Table**: Make a table where you list each student's homework hours and their corresponding grades.
   - **Categorize Data**: Group students into categories based on the amount of time they spend on homework (e.g., 0-1 hours, 1-2 hours, 2-3 hours, etc.).

4. **Analyze the Data**:

    - **Look for Patterns**: Calculate the average grades for each group of students based on their homework hours.
    - **Use Graphs**: Create a graph to visualize the relationship between homework time and grades. This can help you see if there is any trend.

5. **Conclude**:

    - **Interpret the Data**: Based on your analysis, determine if students who spend more time on homework generally have higher grades.
    - **Draw Conclusions**: If a positive trend is observed, you might conclude that more homework time is associated with better grades. If not, you might conclude that other factors influence grades more than homework time.

By identifying the need to understand the relationship between homework time and grades, and collecting and analyzing data, you can draw meaningful conclusions about how students' study habits might affect their academic performance.

➢ **Organization and Display of Data**

- **Using a Circle Graph to Display Data**
  One great way to display data is to use a circle graph.

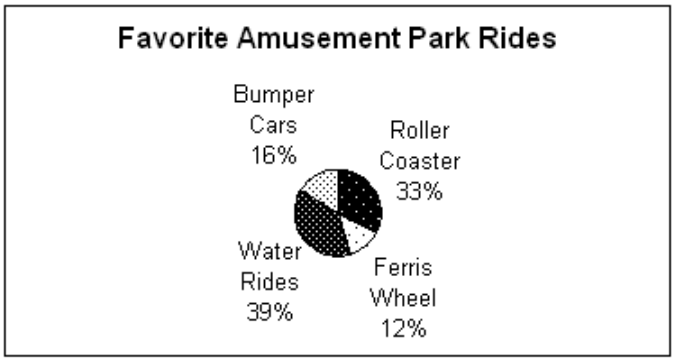
Let's say we polled a group of kids on their favorite amusement park rides. The choices included: • Roller coaster • Ferris wheel • Water rides • Bumper cars	After the survey the percentages were as follows: • Roller coaster: 33% • Ferris wheel: 12% • Water rides: 39% • Bumper cars: 16%

To make a circle graph using this information we must figure out what portion of the circle graph should include each ride. We'll do this by solving proportions. Remember that all circles are 360° and all percents are out of 100%.

Roller Coaster: $\frac{33}{100} = \frac{x}{360}$ $100x = 11880$ $x = 118.8$ We'll round this to 119°.	Ferris Wheel: $\frac{12}{100} = \frac{x}{360}$ $100x = 4320$ $x = 43.2$ We'll round this to 43°.

Water Rides: $\frac{39}{100} = \frac{x}{360}$ $100x = 14040$ $x = 140.4$ We'll round this to 140°.	Bumper Cars: $\frac{16}{100} = \frac{x}{360}$ $100x = 5760$ $x = 57.6$ We'll round this to 58°.

Now, by using a protractor we can divide the circle graph into four pieces using the degree measures that we calculated.



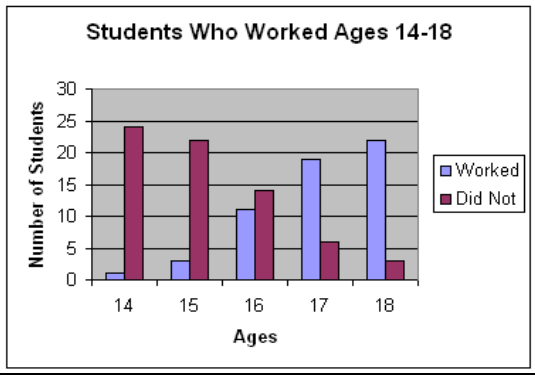
- **Double Bar and Double Line Graphs**

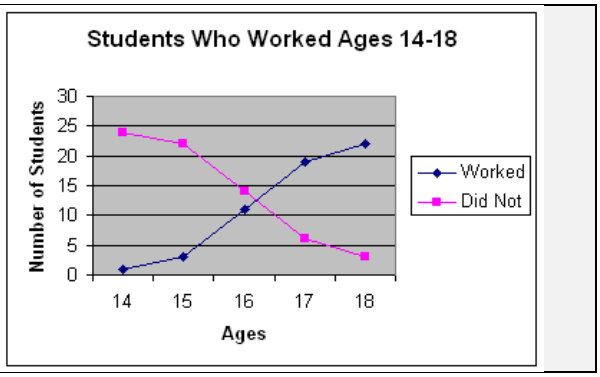
When collecting, organizing, and making sense of raw data, often it is best to make use of bar and line graphs. If comparing two sets of data, we use double bar and double line graphs.

For example, 25 students were followed from age 14 to age 18 to record how many of these students worked part time at each age level. The following is the data that was collected:

14 yrs: 1 worked, 24 did not work
15 yrs: 3 worked, 22 did not work
16 yrs: 11 worked, 14 did not work
17 yrs: 19 worked, 6 did not work
18 yrs: 22 worked, 3 did not work

The next step would be to construct a graph to read the data more easily.

Let's look at a double bar graph Be sure to: • Give the graph a title • Label the axes • Be consistent with the numbers • Use a key • Leave a space between the bars as we go to the next age	

Now let's construct a double line graph using the same data: Be sure to: • Give the graph a title • Label the axes • Be consistent with the numbers • Use a key	

➢ **Analysis of Data**

- **Calculating the Range**

Calculating the **range** for a given set of data is simply finding the difference between the largest number in the data set and the smallest number.

Ex 1: Given the data set of the following test scores:
75, 88, 92, 68, 95, 71, 83
Find the range.

Answer:
The largest number in this data set is 95 while the smallest number is 71. The difference between the two numbers is 24. This is the range of this data set.

- **Measures of Central Tendency**

The measures of central tendency include mean, mode, median and range.

✓ Mean: (same as average) Add the numbers and divide by the number of numbers.
✓ Mode: The number that occurs the most often.
✓ Median: The number is the middle of the data set (after the numbers are ordered least to greatest or vice versa).

Ex 2: Find the mean, median, mode, and range for the following list of values:
13, 18, 13, 14, 13, 16, 14, 21, 13

Answer:
Mean: The mean is the usual average, so:
(13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13) ÷ 9 = 15

Mode: The mode is the number that is repeated more often than any other, so 13 is the mode.
Median: The median is the middle value, so I'll have to rewrite the list in order:
13, 13, 13, 13, 14, 14, 16, 18, 21
There are nine numbers in the list, so the middle one will be the (9 + 1) ÷ 2 = 10 ÷ 2 = 5th number:
13, 13, 13, 13, 14, 14, 16, 18, 21
So, the median is 14.

- **Interpreting Graphs**

When data is represented graphically, we should know how to read and interpret the data.

Data may be represented with pictographs, bar graphs, histograms, line graphs, double line graphs, double bar graphs or circle graphs.

What is being compared in this graph?

Five groups predicted how many m&m's were in a bag then actually counted them. This graph compared the prediction of each group to the actual count.

What is the mean, median, mode and range of both sets of data?

	Prediction	Actual
Mean -	37.4	50
Median -	37	50
Mode -	no mode	52
Range -	25	5

What was the highest prediction and which group made that prediction?
50 m&m's, Group 3

What was the lowest actual count and which group had that count?
47 m&m's, Group 2

➢ **Probability**

Probability is the chance that a given event will occur in an experiment. Fractions are a way to describe the chance that an event will occur.

- **Probability of an Event**

An event, E, is the particular set of outcomes that you're looking for. You can describe the probability of a particular event ($\mathrm{P}(E)$) will happen by using a ratio:

$$\mathrm{P(E)} = \frac{Number\ of\ Equally\ Likely\ Outcomes\ in\ which\ the\ Event\ Can\ Occur}{Total\ Number\ of\ Equally\ Likely\ Possible\ Outcomes}.$$

Ex 3: A spinner has 4 equal sectors colored yellow, blue, green and red. After spinning the spinner, what is the probability of landing on each color?

Answer:

The possible equally likely outcomes of this experiment are yellow, blue, green, and red.

$P(yellow)=\frac{1}{4}$	$P(blue)=\frac{1}{4}$	$P(green)=\frac{1}{4}$	$P(red)=\frac{1}{4}$	

- **Compound Event**

A compound event consists of two or more events. We call probability experiments like these which involve more than one activity "*compound events.*"

The Product Rule:

To find the probability of two events "E" and "F", we can:

$$\mathrm{P(E\ and\ F)} = \mathrm{P(E)} \times \mathrm{P(F\ knowing\ that\ E\ has\ happened).}$$

In the case that the two events are **independent**,

$$\mathrm{P(E\ and\ F)} = \mathrm{P(E)} \times \mathrm{P(F).}$$

Two events are independent if knowing that one event has occurred does not affect the chance of the other event occurring.

Ex 4: A dresser drawer contains two pairs of socks of each of the following colors: blue, brown, red, white, and black. Each pair is folded together in matching pairs.

a) You reach into the sock drawer and choose a pair of socks without looking. The first pair you pull out is red -the wrong color. You replace this pair and choose another pair. What is the probability that you will choose the red pair of socks twice?
b) What would be the probability if we do not replace the first pair of socks selected?

Answer:

In a), we **replace** the pair of socks in the first draw before choosing the next pair of socks in random:

$$\mathrm{P}(Red) = \frac{2}{10} = \frac{1}{5}$$

$$\mathrm{P}(Red\ and\ Red) = \frac{1}{5} \times \frac{1}{5} = \frac{1}{25}.$$

In b), we select the second pair **without replacement**:

$$\mathrm{P}(Red) = \frac{2}{10} = \frac{1}{5}$$

$$\mathrm{P}(Red\ and\ Red) = \frac{2}{10} \times \frac{1}{9} = \frac{2}{90} = \frac{1}{45}.$$