Sampling Techniques

Sampling Techniques

Sampling Techniques

Sampling is the process of selecting a smaller group (sample) from a larger group (population) to collect data and make conclusions about the entire population.

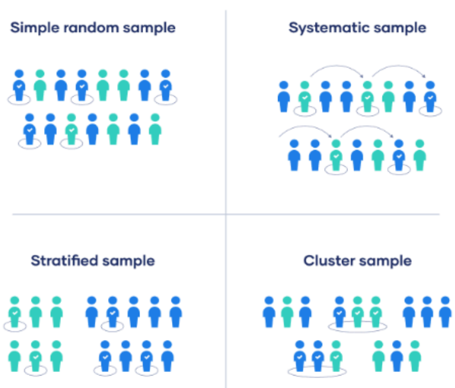
Key Terms:

- **Population:** The entire group you are studying. Example: All students in a school.
- **Sample:** A smaller group selected from the population for analysis. Example: 50 students chosen from the school.
- **Census:** Collecting data from the entire population (very time-consuming and costly).
- **Bias:** A sample that does not fairly represent the population. Example: Asking only basketball players about favorite sports.

Why Use Sampling?

- It is quicker and cheaper than a census.
- Provides useful estimates if the sample is representative.

Main Sampling Methods:



Method	Description
Random Sampling	Each member of the population has an equal chance of selection. Example: Drawing names from a hat.
Systematic Sampling	Choose every nth item from a list after a random start.
Stratified Sampling	Divide population into groups (strata) and select random samples from each proportionally.
Convenience Sampling	Select individuals easiest to reach (may cause bias).

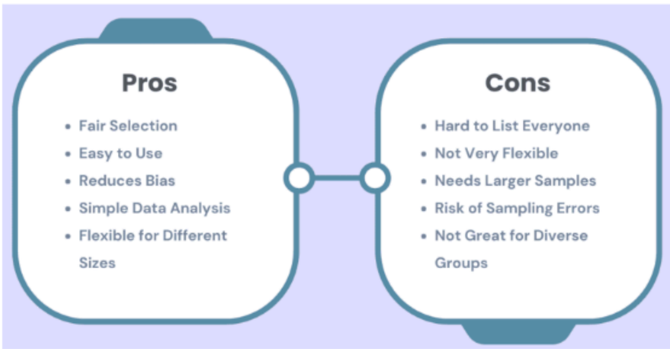
Example:

A researcher selects every 10th household from a list of 1,000 after choosing a random starting point. What sampling method is used?

▼ ▶ **Answer/Explanation**

Answer: This is systematic sampling because items are chosen at regular intervals (every 10th) after a random start.

Advantages of Random Sampling:



- Unbiased if selection is truly random.
- Easy to understand and apply.

Disadvantages:

- Requires a full list of the population.
- Can be time-consuming for large populations.

Data manipulation and misinterpretation

Data manipulation and misinterpretation

Data Manipulation and Misinterpretation

Data manipulation and misinterpretation happen when data is presented or summarized in a way that gives a false or misleading impression. This can occur intentionally (to persuade) or unintentionally (due to poor design or lack of understanding).

Why Misinterpretation Happens:

- **Truncated or inconsistent scales:** Axis does not start at zero or uses uneven intervals, making small changes look significant.
- **Using wrong type of graph:** For example, a pie chart to show change over time.
- **3D effects and distortions:** Can make some data sections appear larger than they are.
- **Missing context:** Lack of information about sample size, source, or time period.
- **Biased sampling:** Non-representative samples leading to incorrect conclusions.

Other Common Issues:

- **Selective reporting:** Showing only part of the data that supports a claim.
- **Misleading averages:** Using mean instead of median when data is skewed.
- **Percentage tricks:** Presenting percentage increases without base values (e.g., "100% increase" from 1 to 2).

How to Avoid Misinterpretation:

- Always check scales and labels on graphs.
- Ask for the sample size and sampling method.
- Choose graphs that correctly represent the data type (e.g., line graph for trends, bar graph for categories).
- Beware of 3D effects or exaggerated colors that draw attention unnecessarily.

Example:

A bar chart compares two sales figures: Year 1 = $98,000 and Year 2 = $102,000. The vertical axis starts at $95,000, making the bars look very different. Why is this misleading?

▼ ▶ Answer/Explanation

Step 1: The actual difference is $4,000, which is small compared to $98,000.

Step 2: Starting the axis at $95,000 exaggerates this difference visually.

Answer: The graph misleads by truncating the axis, making a small increase seem large.

Example 2:

A company advertises "Sales increased by 200%!" without revealing that sales went from 5 units to 15 units. Why is this misleading?

▼ ▶ Answer/Explanation

Step 1: Percentage looks large but actual increase is small (10 units).

Step 2: Without original values, the statement can mislead customers into thinking growth is huge.

Answer: Always check both percentage and absolute values before interpreting growth.

Example:

A graph compares two schools' average test scores. School A = 68% School B = 72%. The bars are drawn with a big gap, and the scale starts at 60%. What impression does this give?

▼ **Answer/Explanation**

Step 1: Actual difference is 4%, very small.

Step 2: Because the axis starts at 60%, the gap looks very large.

Answer: The graph is misleading because of a truncated scale exaggerating a small difference.

Important Notes:

- Percentages in a pie chart must add up to 100%.
- Always look for axis starting points and scale increments.
- Ask for actual numbers, not just percentages.
- Beware of terms like "doubles," "triples," or "huge increase" without context.

Graphical Representations

Graphical Representations

Graphical representation of data is the process of presenting numerical data visually using charts or graphs. This helps identify patterns, trends, and relationships quickly.

Why use Graphs?

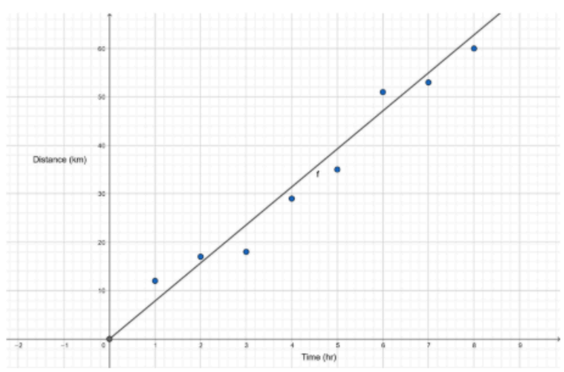
- Makes data easier to understand and interpret.
- Helps identify trends and relationships between variables.
- Used for comparison and decision-making.

Main Types of Graphs in MYP:

- **Bivariate Graphs:** Show the relationship between two variables (e.g., time vs distance).
- **Scatter Graphs:** Show how two continuous variables relate and can indicate correlation.
- **Box Plots (Box-and-Whisker):** Show distribution using median, quartiles, and outliers.
- **Cumulative Frequency Graphs:** Show how cumulative totals build up, useful for medians and percentiles.

Bivariate Graphs

A bivariate graph shows the relationship between two variables plotted on a coordinate plane. One variable is placed on the x-axis (independent) and the other on the y-axis (dependent).



Key Features:

- Helps visualize the connection between two quantities.
- Usually plotted as points or a line graph when the data is continuous.
- Common in real-life: time vs speed, age vs height.

Example:

The table shows time (minutes) and distance (km) traveled by a cyclist:

Time (minutes)	Distance (km)
0	0
10	5
20	10
30	15
40	20

Draw a bivariate graph and describe the relationship.

▼ ▶ Answer/Explanation

Step 1: Plot time on the x-axis and distance on the y-axis.

Step 2: Plot points (0,0), (10,5), (20,10), (30,15), (40,20).

Step 3: Join with a straight line – this shows a linear relationship.

Answer: The graph shows direct proportionality: as time increases, distance increases at a constant rate.

Scatter Graphs

Scatter Graphs

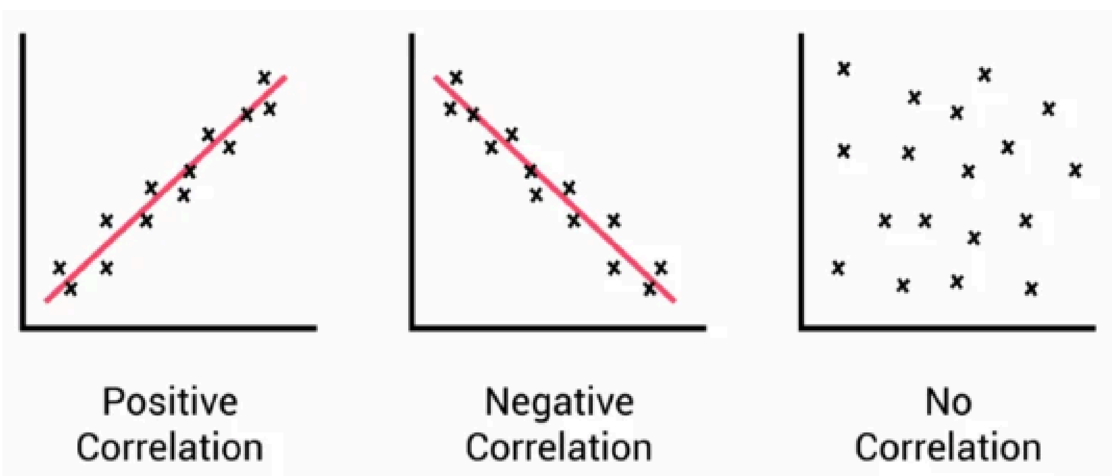
A scatter graph (or scatter plot) shows the relationship between two numerical variables by plotting points on a coordinate plane. It is used to check for correlation between variables.

Key Features:

- Each point represents one pair of values (x, y).
- Used to check for a **correlation** between two variables.
- We can draw a **line of best fit** to show the trend.

Types of Correlation:

- **Positive correlation:** As one variable increases, the other also increases.
- **Negative correlation:** As one variable increases, the other decreases.
- **No correlation:** No pattern or relationship between the two variables.



.ine of Best Fit:

- A straight line drawn through the points, showing the overall trend.
- Helps predict values of one variable based on the other.

Example:

The table shows the number of hours studied and the marks scored by 8 students:

Hours Studied	Marks (%)
1	45
2	50
3	58
4	62
5	70
6	74
7	78
8	85

Draw a scatter graph and describe the correlation.

▼ ▶ **Answer/Explanation**

Step 1: Plot points (Hours, Marks) on a graph.

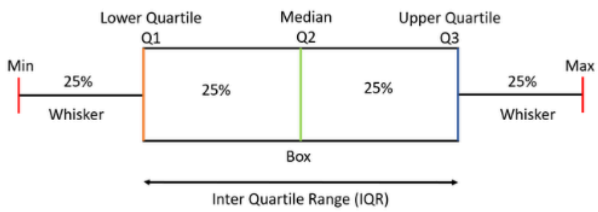
Step 2: Observe the pattern: points rise upwards from left to right.

Step 3: Draw a line of best fit through the data points.

Answer: There is a strong positive correlation between hours studied and marks scored.

Box Plots (Box-and-Whisker)

A box plot is a diagram that shows the distribution of a data set using five key values: minimum, lower quartile (Q1), median (Q2), upper quartile (Q3), and maximum. It helps visualize the spread and detect outliers.



Key Features:

- The box represents the interquartile range (IQR) = Q3 − Q1.
- A line inside the box marks the median.
- Whiskers extend to the smallest and largest data values.
- Shows symmetry, skewness, and spread of data.

Steps to Draw a Box Plot:

1. Arrange data in ascending order.
2. Find:
   - **Minimum** and **Maximum**
   - **Median (Q2)**
   - **Lower quartile (Q1):** median of the lower half
   - **Upper quartile (Q3):** median of the upper half
3. Draw a scale and plot these five values.
4. Draw the box from Q1 to Q3 and whiskers to min and max.

Example:

The marks (out of 50) scored by 9 students are:

Marks
12
15
18
20
24
26
28
32
36

Draw a box plot for this data.

▼ ▶ **Answer/Explanation**

Step 1: Arrange data (already sorted): 12, 15, 18, 20, 24, 26, 28, 32, 36.

Step 2: Find five-number summary:

Min = 12

Q_1 = 18 (middle of first 4 values)

Median (Q_2) = 24

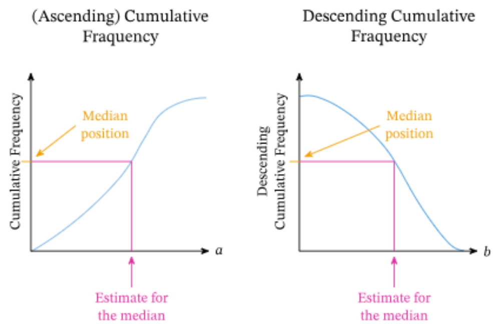
Q_3 = 28 (middle of last 4 values)

Max = 36

Step 3: Draw box from 18 to 28, line at 24, whiskers at 12 and 36.

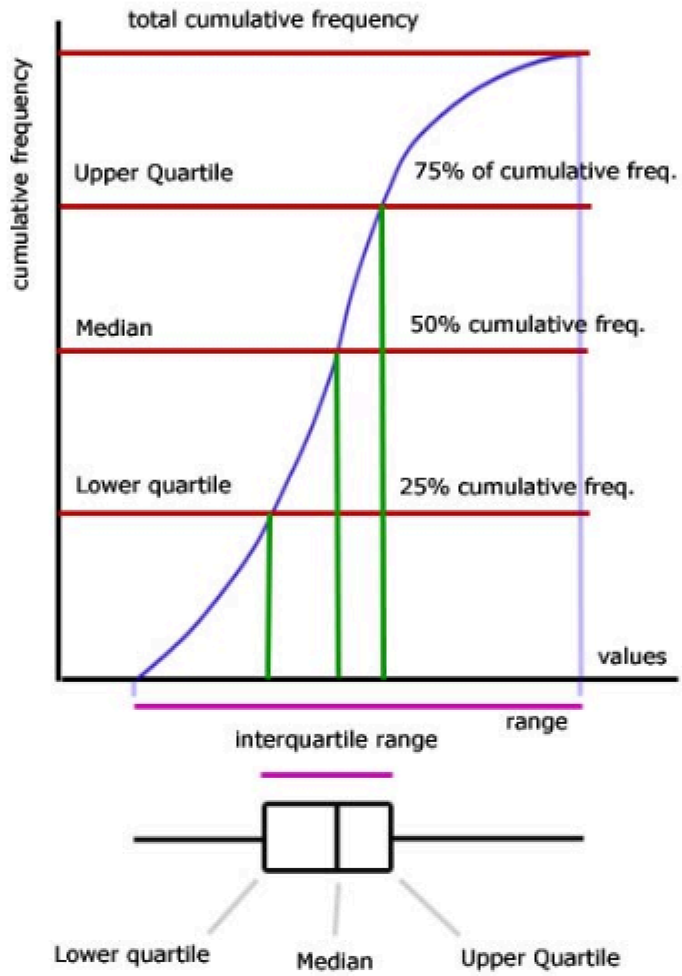
Cumulative Frequency Graphs

A cumulative frequency graph shows how the cumulative total of a data set increases as the variable increases. It is useful for estimating medians, quartiles, and percentiles.



Key Features:

- The cumulative frequency is the running total of frequencies up to a certain point.
- The graph is usually drawn as an **ogive** (a smooth curve or straight lines connecting points).
- The horizontal axis shows the upper class boundaries, and the vertical axis shows cumulative frequency.



Steps to Draw a Cumulative Frequency Graph:

1. Create a cumulative frequency table from the given data.
2. Plot cumulative frequency against upper class boundary for each class.
3. Draw a smooth curve or join with straight lines.
4. Use the graph to estimate median, quartiles, and percentiles.

Example:

The weights (in kg) of 50 students are given below:

Weight (kg)	Frequency
40–50	5
50–60	10
60–70	20
70–80	10
80–90	5

Draw the cumulative frequency graph and find the interquartile range (IQR).

▼ ▶ **Answer/Explanation**

Step 1: Calculate cumulative frequency:

Upper Boundary	Cumulative Frequency
50	5
60	15
70	35
80	45
90	50

Step 2: Plot the graph and read quartiles:

Q_1 = 25% of 50 = 12.5th value ≈ 58 kg

Q_3 = 75% of 50 = 37.5th value ≈ 72 kg

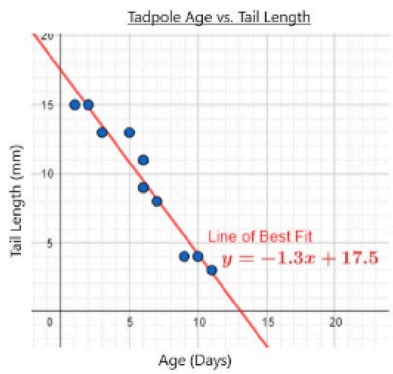
Step 3: IQR = $Q_3 - Q_1 = 72 - 58 = 14$ kg.

Lines of Best Fit

Lines of Best Fit

Lines of Best Fit

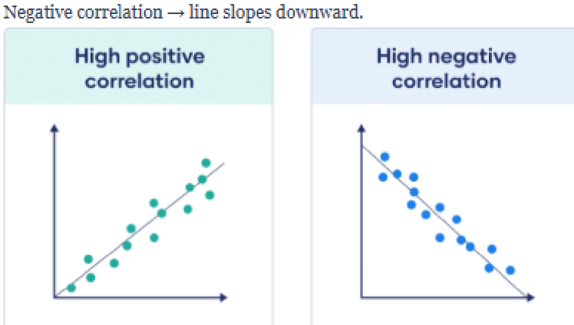
A line of best fit is a straight line drawn through a scatter plot to show the general trend of the data. It is used to make predictions and analyze relationships between variables.



Key Features:

- The line passes as close as possible to all points (not all points will lie on the line).
- Used to estimate values that are not in the original data (interpolation and extrapolation).
- The line should have roughly equal points above and below it.

Steps to Draw a Line of Best Fit:

1. Plot the scatter graph for the data set.
2. Observe the correlation:
   - Positive correlation → line slopes upward.
   - Negative correlation → line slopes downward.
   - 

3. Draw a straight line with approximately equal points on both sides.
4. Use the line to predict values.

Example:

The table shows the number of study hours and test scores for 6 students:

Hours	Score (%)
1	40
2	48
3	56
4	63
5	70
6	78

Draw a line of best fit and predict the score for 7 hours of study.

▼ **Answer/Explanation**

Step 1: Plot points (Hours, Score) on a scatter graph.

Step 2: Draw a line of best fit through the points with equal balance on both sides.

Step 3: Extend the line to Hours = 7 and read Score ≈ 85%.

Answer: The predicted score for 7 hours is about 85%.

Example:

The table shows the outside temperature and the number of jackets sold:

Temperature (°C)	Jackets Sold
5	50
8	45
12	35
15	28
18	22
20	18

Draw the line of best fit and predict jackets sold at 10°C.

▼ ▶ Answer/Explanation

Step 1: Plot points (Temperature, Jackets sold).

Step 2: Draw the line of best fit with roughly equal points on both sides.

Step 3: At 10°C, read from the line ≈ 40 jackets.

Answer: Predicted sales at 10°C = 40 jackets.

3. Draw a straight line with approximately equal points on both sides.
4. Use the line to predict values.

Example:

The table shows the number of study hours and test scores for 6 students:

Hours	Score (%)
1	40
2	48
3	56
4	63
5	70
6	78

Draw a line of best fit and predict the score for 7 hours of study.

▼ ▶ **Answer/Explanation**

Step 1: Plot points (Hours, Score) on a scatter graph.

Step 2: Draw a line of best fit through the points with equal balance on both sides.

Step 3: Extend the line to Hours = 7 and read Score ≈ 85%.

Answer: The predicted score for 7 hours is about 85%.

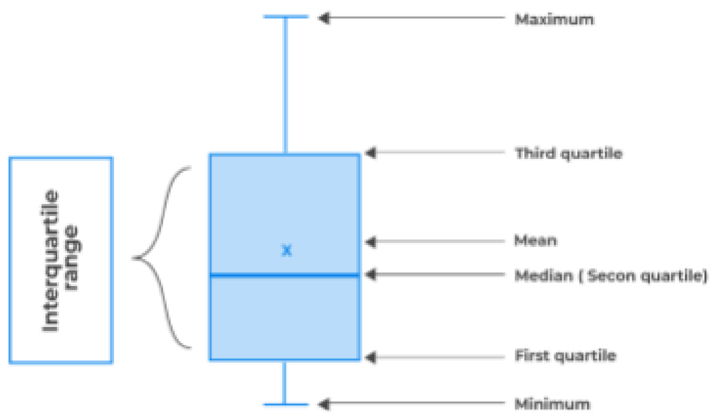
Data processing: quartiles and percentiles

Data processing: quartiles and percentiles

Quartiles

Quartiles divide a set of data into four equal parts after arranging the data in ascending order:

- **Q1 (Lower Quartile):** 25% of the data ($\frac{1}{4}$ position).
- **Q2 (Median):** 50% of the data (middle value).
- **Q3 (Upper Quartile):** 75% of the data ($\frac{3}{4}$ position).



Interquartile Range (IQR): $Q3 - Q1$, measures the spread of the middle 50% of data.

Steps to Calculate Quartiles:

1. Arrange data in ascending order.
2. Find positions:
   - $Q1$ at position $\frac{n+1}{4}$
   - $Q2$ at position $\frac{n+1}{2}$
   - $Q3$ at position $\frac{3(n+1)}{4}$
3. If the position is not an integer, interpolate between two values.

Example:

Find Q1, Q2, and Q3 for the data: 6, 8, 10, 12, 14, 18, 20.

▼ ▶ **Answer/Explanation**

Step 1: Data in order: 6, 8, 10, 12, 14, 18, 20 ($n = 7$).

Step 2: Positions:
Q1 at $\frac{7+1}{4} = 2 \rightarrow$ 2nd value = 8
Q2 at $\frac{7+1}{2} = 4 \rightarrow$ 4th value = 12
Q3 at $\frac{3(7+1)}{4} = 6 \rightarrow$ 6th value = 18

Answer: Q1 = 8, Q2 = 12, Q3 = 18.

Example:

The weights (kg) of 8 students: 40, 42, 44, 46, 48, 52, 54, 60. Find Q1, Q2, and Q3.

▼ ▶ **Answer/Explanation**

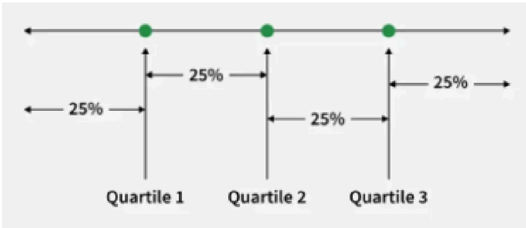
Step 1: Data sorted: 40, 42, 44, 46, 48, 52, 54, 60 ($n = 8$).

Step 2: Positions:
Q1 at $\frac{8+1}{4} = 2.25 \rightarrow$ between 2nd (42) and 3rd (44): $Q1 = 42 + 0.25(44 - 42) = 42.5$
Q2 at $\frac{8+1}{2} = 4.5 \rightarrow$ between 4th (46) and 5th (48): $Q2 = 46 + 0.5(48 - 46) = 47$
Q3 at $\frac{3(8+1)}{4} = 6.75 \rightarrow$ between 6th (52) and 7th (54): $Q3 = 52 + 0.75(54 - 52) = 53.5$

Answer: Q1 = 42.5, Q2 = 47, Q3 = 53.5.

Percentiles

Percentiles divide a data set into 100 equal parts. The k^{th} percentile (P_k) is the value below which k% of the data falls.



- P_{25} = 25th percentile (same as Q1).
- P_{50} = 50th percentile (same as the median).
- P_{75} = 75th percentile (same as Q3).

Steps to Calculate Percentiles:

1. Arrange the data in ascending order.
2. Find position of P_k: $\text{Position} = \frac{k}{100}(n+1)$
3. If the position is not an integer, interpolate between values.

Example:

The test scores are: 12, 18, 20, 25, 28, 30, 35, 40, 42, 50. Find the 70th percentile (P_{70}).

▼ Answer/Explanation

Step 1: Arrange data (already sorted), $n = 10$.

Step 2: Position = $\frac{70}{100}(10+1) = 0.7 \times 11 = 7.7$.

7th value = 35, 8th value = 40.

Interpolate: $35 + 0.7(40–35) = 35 + 3.5 = 38.5$.

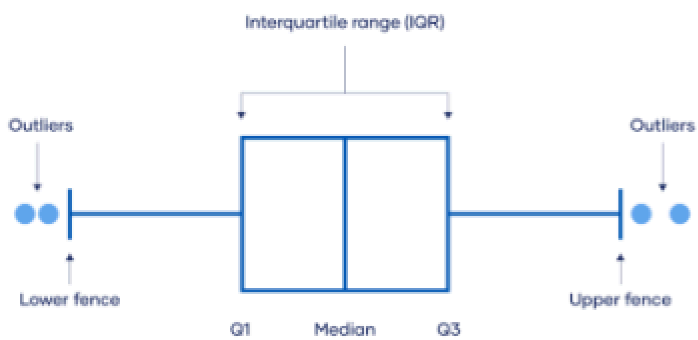
Answer: 70th percentile = 38.5.

Interquartile Range (IQR)

Interquartile Range (IQR)

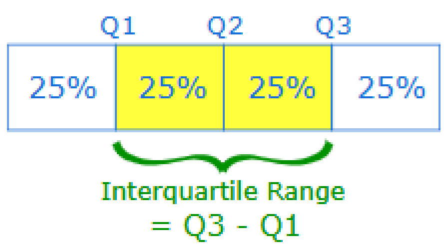
Interquartile Range (IQR)

The Interquartile Range (IQR) measures the spread of the middle 50% of data.



It is calculated as:

$\text{IQR} = Q_3 – Q_1$

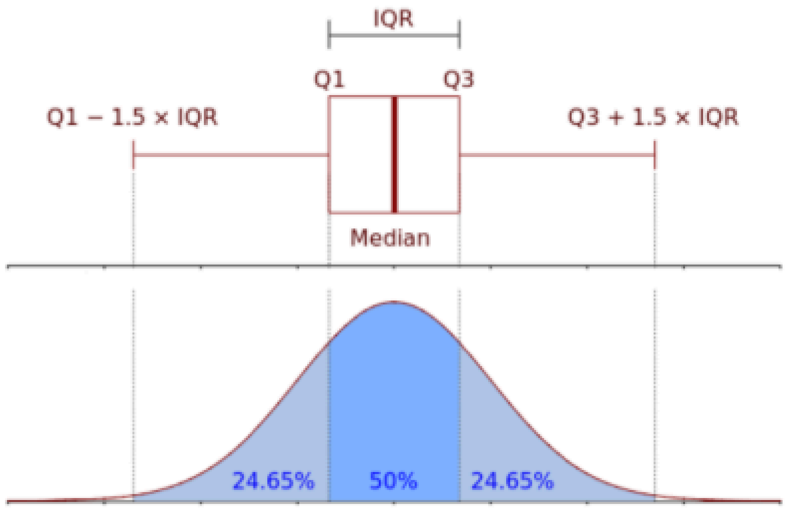


Why use IQR?

- Not affected by outliers, unlike the full range.
- Shows how tightly data is clustered around the median.

Steps to Calculate IQR:

1. Arrange the data in ascending order.
2. Find Q_1 (lower quartile) and Q_3 (upper quartile).
3. Compute $IQR = Q_3 – Q_1$.



Example:

Find the IQR for the data: 8, 10, 12, 14, 16, 18, 20, 22.

▼ **Answer/Explanation**

Step 1: Data is sorted ($n = 8$).

Step 2:

Q_1 position = $\frac{8+1}{4} = 2.25 \rightarrow$ between 2nd (10) and 3rd (12): $Q_1 = 10 + 0.25(12 - 10) = 10.5$

Q_3 position = $\frac{3(8+1)}{4} = 6.75 \rightarrow$ between 6th (18) and 7th (20): $Q_3 = 18 + 0.75(20 - 18) = 19.5$

Step 3: IQR = $19.5 – 10.5 = 9$.

Answer: IQR = 9.

Example:

The marks of 9 students are: 15, 18, 21, 24, 26, 30, 32, 35, 40. Calculate the IQR.

▼ Answer/Explanation

Step 1: Sorted data: (already sorted), $n = 9$.

Step 2:

Q_1 position = $\frac{9+1}{4} = 2.5 \rightarrow$ between 2nd (18) and 3rd (21): $Q_1 = 18 + 0.5(21 - 18) = 19.5$

Q_3 position = $\frac{3(9+1)}{4} = 7.5 \rightarrow$ between 7th (32) and 8th (35): $Q_3 = 32 + 0.5(35 - 32) = 33.5$

Step 3: IQR = $33.5–19.5 = 14$.

Answer: IQR = 14.

Example:

The times (in minutes) taken by 10 students to complete an online quiz: 12, 14, 15, 18, 20, 21, 23, 25, 27, 30. Calculate the IQR.

▼ Answer/Explanation

Step 1: Data is sorted ($n = 10$).

Step 2:

Q_1 position = $\frac{10+1}{4} = 2.75 \rightarrow$ between 2nd (14) and 3rd (15): $Q_1 = 14 + 0.75(15 - 14) = 14.75$

Q_3 position = $\frac{3(10+1)}{4} = 8.25 \rightarrow$ between 8th (25) and 9th (27): $Q_3 = 25 + 0.25(27 - 25) = 25.5$

Step 3: IQR = $25.5–14.75 = 10.75$.

Answer: IQR = 10.75 minutes.

Correlation and Qualitative Data Handling

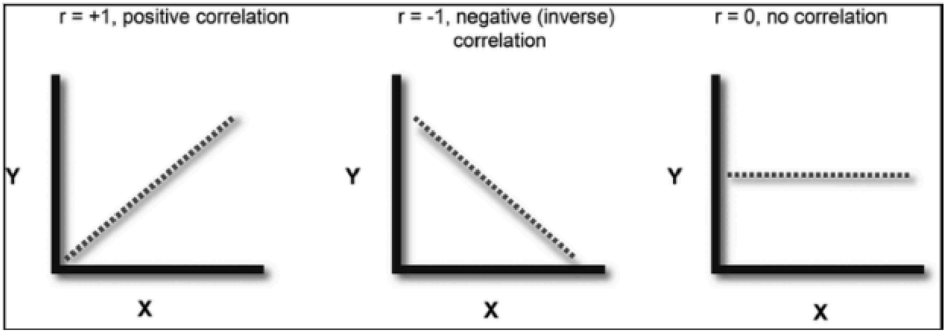
Correlation and Qualitative Data Handling

Correlation

Correlation is a statistical measure that describes the strength and direction of a relationship between two variables.

Types of Correlation:

- **Positive Correlation:** As one variable increases, the other also increases (e.g., height and weight).
- **Negative Correlation:** As one variable increases, the other decreases (e.g., speed and travel time).
- **No Correlation:** No predictable relationship between variables.



Methods to Represent Correlation:

- Scatter graphs with a line of best fit.
- Correlation coefficient (r) values from -1 to +1:
  - $r \approx +1$: Strong positive correlation
  - $r \approx -1$: Strong negative correlation
  - $r \approx 0$: No correlation

Example:

The table shows the hours studied and marks scored by 6 students:

Hours Studied	Marks Scored
2	40
3	50
4	55
5	65
6	70
7	80

What type of correlation exists between hours studied and marks scored?

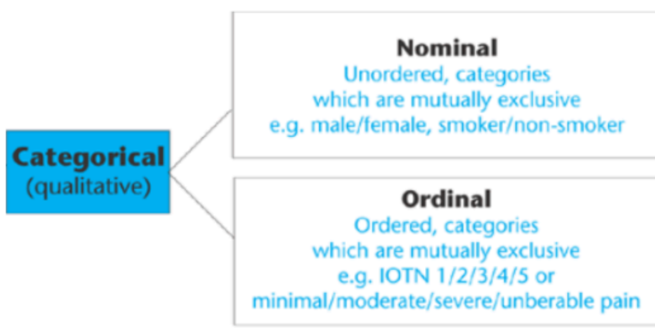
▼ Answer/Explanation

Observation: As hours studied increase, marks scored also increase.

Conclusion: Positive correlation exists between hours studied and marks scored.

Qualitative Data Handling

Qualitative data refers to non-numeric information describing categories or attributes (e.g., colors, brands, opinions).



It can be classified into:

- **Nominal:** Categories with no natural order (e.g., blood group, eye color).
- **Ordinal:** Categories with an order but no fixed difference (e.g., satisfaction: Poor, Fair, Good).

Methods to Handle Qualitative Data:

1. **Frequency Table:** Lists each category and its frequency (number of occurrences). Useful for summarizing raw data.
2. **Bar Chart:** Uses rectangular bars with lengths proportional to frequencies. Gaps between bars indicate categories are discrete.
3. **Pie Chart:** Represents data as sectors of a circle. Each sector's angle = $\frac{\text{frequency}}{\text{total}} \times 360^\circ$.
4. **Pictogram:** Uses pictures or symbols to represent frequency. Each symbol represents a fixed number of items.

Example :

20 students were asked their favorite sport. The results were:

Football, Basketball, Football, Cricket, Cricket, Football, Basketball, Football, Tennis, Cricket, Football, Basketball, Tennis, Football, Cricket, Football, Basketball, Football, Tennis, Cricket.

▼ **Answer/Explanation**

Step 1: Count each category:

Sport	Frequency
Football	8
Basketball	4
Cricket	5
Tennis	3

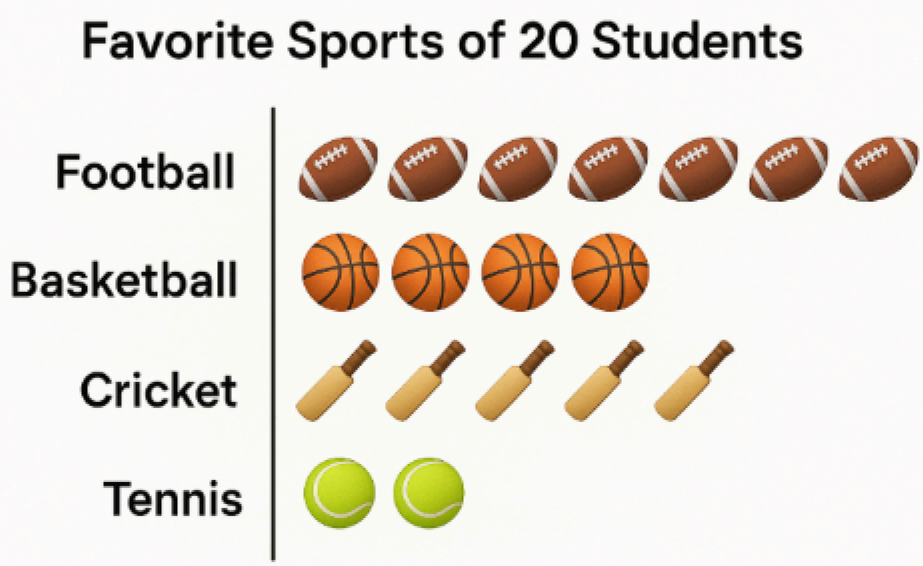
Example :

Represent the same data as a pictogram where one Sports symbol = 1 students.

▼ **Answer/Explanation**

Symbols required:

- Football: 8 ÷ 1 = 8 symbols
- Basketball: 4 ÷ 1 = 4 symbols
- Cricket: 5 ÷ 1 = 5 symbols
- Tennis: 3 ÷ 1 = 3 symbols



Relative and Expected Frequency

Relative and Expected Frequency

Relative Frequency

Relative frequency is the ratio of the number of times an event occurs to the total number of trials. It estimates the probability of an event based on experimental data.

Formula

$$\text{Relative Frequency} = \frac{\text{Subgroup Count}}{\text{Total Count}} = \frac{f}{n}$$

$$\text{Relative Frequency} = \frac{\text{Number of successful outcomes}}{\text{Total number of trials}}$$

Key Points:

- The more trials conducted, the closer the relative frequency comes to the theoretical probability (Law of Large Numbers).
- Used when theoretical probability is hard to calculate.

Example:

A coin was flipped 100 times and landed on heads 58 times. Find the relative frequency of getting heads.

▼ **Answer/Explanation**

Step 1: Number of successful outcomes = 58, Total trials = 100.

Step 2: Relative frequency = $\frac{58}{100} = 0.58$.

Interpretation: The experimental probability of getting heads is approximately 0.58 (or 58%).

Expected Frequency

Expected frequency is the number of times an event is likely to occur based on its theoretical probability in a given number of trials.

$$\text{Expected Frequency} = \text{Total Trials} \times \text{Theoretical Probability}$$

Key Points:

- Expected frequency is a prediction based on probability, not actual results.
- It is useful in probability experiments and statistical tests (e.g., Chi-square test).

Example:

A fair coin is tossed 200 times. What is the expected frequency of getting heads?

▼ Answer/Explanation

Step 1: Theoretical probability of heads = $\frac{1}{2} = 0.5$.

Step 2: Total trials = 200.

Step 3: Expected frequency = $200 \times 0.5 = 100$.

Answer: We expect 100 heads in 200 tosses.

Example:

A die is rolled 600 times. What is the expected frequency of getting an even number (2, 4, or 6)?

▼ Answer/Explanation

Step 1: Theoretical probability of an even number = $\frac{3}{6} = 0.5$.

Step 2: Total trials = 600.

Step 3: Expected frequency = $600 \times 0.5 = 300$.

Answer: We expect 300 even numbers in 600 rolls.

Response Rates

Response Rates

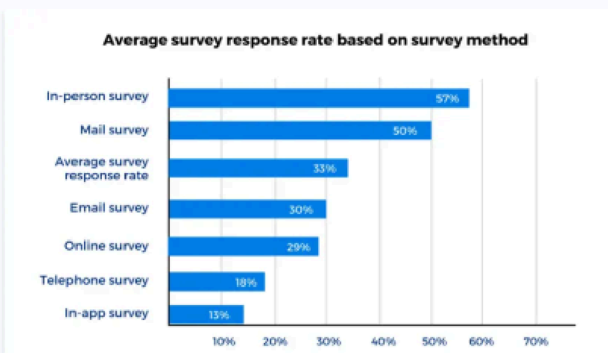
Response Rates

Response rate refers to the proportion of people who respond to a survey or questionnaire out of the total number of people contacted. It is an important measure of the quality and reliability of survey data.

$$\text{Survey response rate} = \frac{\text{Number of people who participated in the survey}}{\text{Total number of recipients you sent it to}} \times 100$$

$$\text{Response Rate} = \frac{\text{Number of Responses}}{\text{Number of People Contacted}} \times 100\%$$

Key Points:

- A higher response rate generally indicates more reliable data.
    - 

- Low response rates can lead to **non-response bias**, making results less representative.
- Methods to improve response rate:
    - Sending reminders
    - Keeping surveys short and simple
    - Offering incentives

Example:

A company sends out 500 feedback forms and receives 400 responses. Find the response rate.

▼ **Answer/Explanation**

Step 1: Number of responses = 400, Number contacted = 500.

Step 2: Response rate = $\frac{400}{500} \times 100\% = 80\%$.

Answer: The response rate is 80%.

Example:

A teacher emails a survey to 120 students, but only 72 reply. Calculate the response rate and suggest one way to improve it.

▼ **Answer/Explanation**

Step 1: Number of responses = 72, Number contacted = 120.

Step 2: Response rate = $\frac{72}{120} \times 100\% = 60\%$.

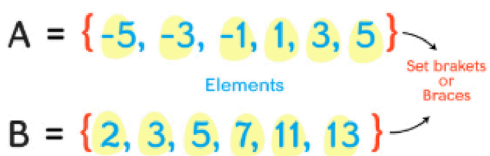
Suggestion: Send reminders or offer a small incentive to increase participation.

Sets: Definition and Notation

Sets: Definition and Notation

Sets: Definition and Representation

A set is a well-defined collection of distinct objects, called elements. Sets are represented by capital letters and their elements by curly brackets.

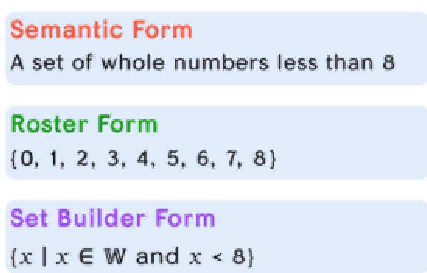


Examples: $A = \{1, 2, 3\},\ B = \{\text{apple}, \text{banana}, \text{grape}\}$

Elements and Cardinality

- An **element** of a set is an individual object in that set.
- **Notation:** $x \in A$ means "x is an element of A".
- **Cardinality:** The number of elements in a set. Denoted by $n(A)$.
- **Example:** For $A = \{1, 2, 3, 4\}$, $n(A) = 4$.

Representing Sets



- **Roster/Tabular Form:** List elements inside curly braces. Example: $A = \{2, 4, 6, 8\}$
- **Set-Builder Form:** Describe properties of elements. Example: $A = \{x : x \text{ is an even number } \leq 8\}$
- **Venn Diagram:** Use circles inside a rectangle (universal set) to represent relationships among sets.

Types of Sets

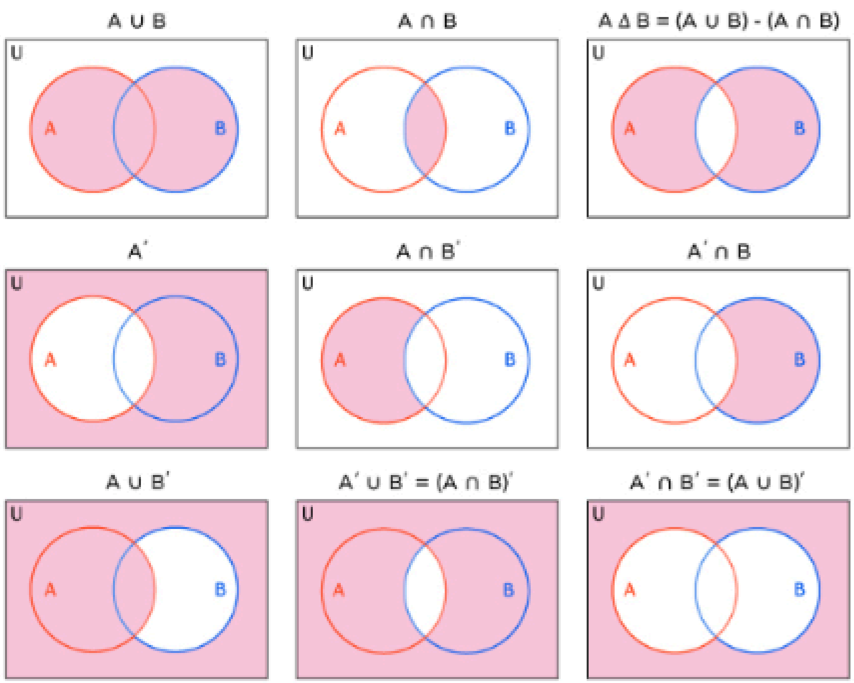
Type	Description	Example
Empty Set (∅)	No elements	$\{\}$
Finite Set	Countable number of elements	$\{1, 2, 3\}$
Infinite Set	Uncountable elements	$\{1, 2, 3, \ldots\}$
Equal Sets	Same elements	$\{a, b\} = \{b, a\}$
Subset	Every element of A is in B	$A \subset B$

Common Set Notation:

Notation	Meaning
$\in$	Element of
$\notin$	Not an element of
$\subset$	Subset of
$\cup$	Union
$\cap$	Intersection
U	Universal Set
$\emptyset$	Empty Set

Venn Diagram Representation:

- Universal set U is represented by a rectangle.
- Each set (A, B, C) is shown as a circle inside the rectangle.
- $A \cup B$: Shaded area covering both circles (Union).
- $A \cap B$: Overlapping region of two circles (Intersection).
- $A – B$: Area in A but outside B (Difference).
- A': Area outside A but inside the rectangle (Complement).



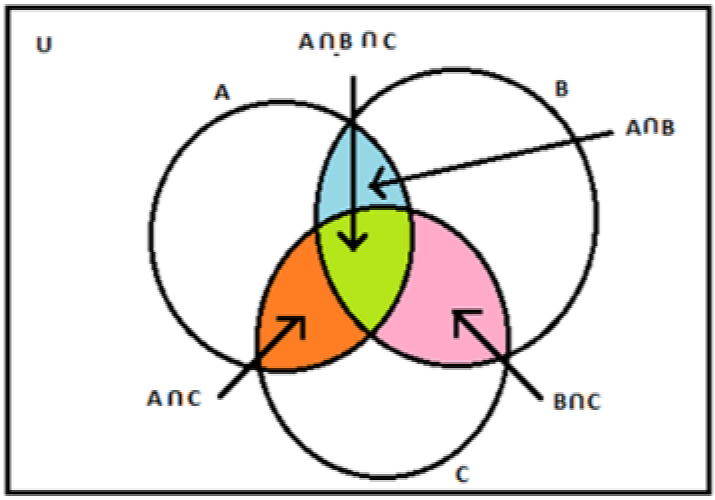
Operations on Sets:

- **Union ($A \cup B$):** All elements in A or B or both.
- **Intersection ($A \cap B$):** All elements common to A and B.
- **Difference ($A – B$):** Elements in A but not in B.
- **Complement (A'):** All elements in universal set U but not in A.

For Three Sets (A, B, C):

- $A \cup B \cup C$: All elements in A or B or C.
- $A \cap B \cap C$: Elements common to all three sets.
- $(A \cup B) \cap C$: Elements in C and at least one of A or B.

Venn Diagram Representation:



Properties of Sets

- **Commutative:** $A \cup B = B \cup A,\ A \cap B = B \cap A$
- **Associative:** $A \cup (B \cup C) = (A \cup B) \cup C$
- **Distributive:** $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- **Identity:** $A \cup \emptyset = A,\ A \cap U = A$
- **Complement Laws:** $A \cup A' = U,\ A \cap A' = \emptyset$

Example:

If A and B are two sets such that $n(A) = 12$, $n(B) = 8$, and $n(A \cup B) = 15$, find $n(A \cap B)$.

▼ Answer/Explanation

Step 1: Use the formula: $n(A \cup B) = n(A) + n(B) – n(A \cap B)$

Step 2: Substitute values: $15 = 12 + 8 – n(A \cap B)$

Step 3: $n(A \cap B) = 12 + 8 – 15 = 5$

Answer: $n(A \cap B) = 5$.

Example:

If $A = \{2, 4, 6, 8\}$, $B = \{4, 6, 8, 10, 12\}$, and the universal set $U = \{2, 4, 6, 8, 10, 12, 14\}$, find:

(a) $A \cup B$,

(b) $A \cap B$,

(c) A' (complement of A).

▼ ▶ **Answer/Explanation**

(a) $A \cup B = \{2, 4, 6, 8, 10, 12\}$

(b) $A \cap B = \{4, 6, 8\}$

(c) $A' = \{10, 12, 14\}$ (elements in U but not in A).

Example:

Express the set $A = \{3, 6, 9, 12, 15, 18\}$ in set-builder notation.

▼ ▶ **Answer/Explanation**

Observation: These numbers are multiples of 3 up to 18.

So, $A = \{x : x = 3n,\ 1 \leq n \leq 6,\ n \in \mathbb{N}\}$.

Example:

In a survey of 100 students:
• 50 like tea (T)
• 40 like coffee (C)
• 20 like both

(a) Represent this using a Venn diagram.

(b) How many like only tea?

▼ **Answer/Explanation**
Step 1: Use two-set Venn diagram with overlap = 20.

Tea only: $50 – 20 = 30.$

Coffee only: $40 – 20 = 20.$

Answer: 30 students like only tea.

Example:

In a class:

- 10 study Math (M)
- 8 study Science (S)
- 6 study English (E)
- 3 study all three subjects
- 2 study Math and Science only
- 1 studies Science and English only

How many study only Math?

▼ ▶ **Answer/Explanation**

Step 1: Start with intersection of all three = 3.

Math and Science only: 2 (already includes all three, so add separately).

Total in Math: 10 = only Math + 2 + 3.

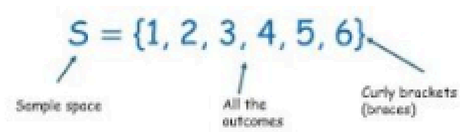
So only Math: 10 – (2 + 3) = 5.

Probability venn diagram

Probability with Venn diagrams, tree diagrams and sample spaces

Sample Space

The sample space of an experiment is the set of all possible outcomes. It is often denoted by S.



Key Points:

- Each outcome is called a sample point.
- Probability of an event = $\dfrac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}}$.
- Sample space can be represented by a list, table, or diagram.

Example:

Two dice are rolled. Construct the sample space and find the probability of getting:

- (a) A sum of 7
- (b) A double (same numbers)

▼ **Answer/Explanation**

Step 1: Sample space size = $6 \times 6 = 36$.

Step 2: Sum of 7: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) → 6 outcomes.

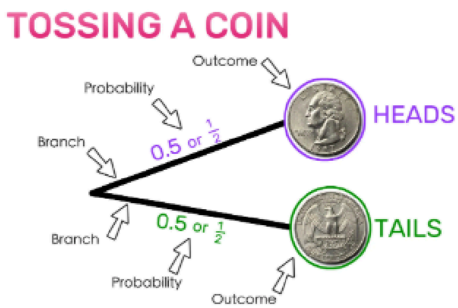
$$P(\text{sum of } 7) = \frac{6}{36} = \frac{1}{6}.$$

Step 3: Doubles: (1,1), (2,2), (3,3), (4,4), (5,5), (6,6) → 6 outcomes.

$$P(\text{double}) = \frac{6}{36} = \frac{1}{6}.$$

Tree Diagrams in Probability

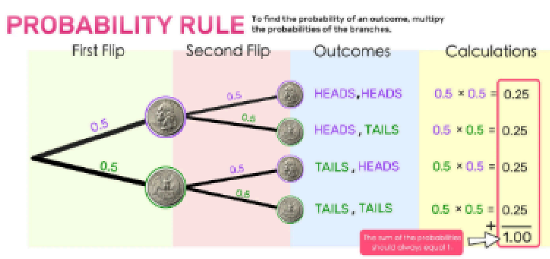
A tree diagram is a visual representation of all possible outcomes of an event and their associated probabilities. It is especially useful for compound events and calculating probabilities using the multiplication rule.



Key Points:

- Each branch represents an outcome and its probability.
- The sum of probabilities on branches from a single node = 1.
- To find the probability of a combined event, multiply along the branches.

Multiplication Rule: $P(A \text{ and } B) = P(A) \times P(B \mid A)$



Example:

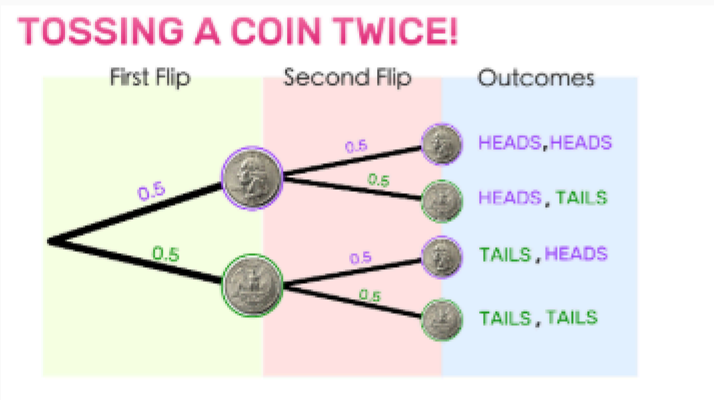
A coin is tossed twice. Use a tree diagram to find the probability of getting:

- (a) Two heads
- (b) One head and one tail

▼ **Answer/Explanation**

Step 1: Outcomes for first toss: H, T. Outcomes for second toss: H, T.

Step 2: Probabilities for each branch = 0.5.



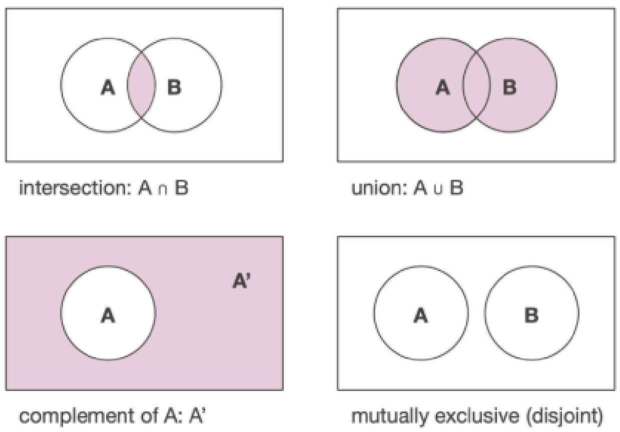
Step 3:
(a) Two heads = $0.5 \times 0.5 = 0.25.$
(b) One head, one tail = HT or TH = $0.25 + 0.25 = 0.5.$

Venn Diagrams in Probability

Venn diagrams visually represent sets and their relationships. In probability, they are used to show sample spaces and calculate probabilities of events, including unions, intersections, and complements.

Key Concepts:

- **Union:** $P(A \cup B) = P(A) + P(B) – P(A \cap B)$
- **Intersection:** $P(A \cap B)$ = Probability that both events occur.
- **Complement:** $P(A') = 1 – P(A)$
- **Mutually Exclusive:** If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.



Example:

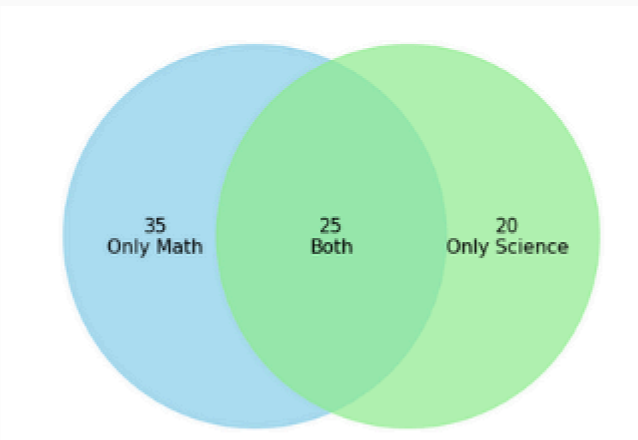
In a group of 100 students:
• 60 like Mathematics (M)
• 45 like Science (S)
• 25 like both

(a) Draw a Venn diagram. (b) Find the probability that a student likes:

- (i) Both subjects
- (ii) Only Mathematics
- (iii) At least one subject

▼ **Answer/Explanation**

Step 1: Use the formula:



Only Math = $60 – 25 = 35$

Only Science = $45 – 25 = 20$

Both = 25

None = $100 – (35 + 20 + 25) = 20$

Step 2: Probabilities:

- (i) $P(\text{Both}) = \frac{25}{100} = 0.25$
- (ii) $P(\text{Only Math}) = \frac{35}{100} = 0.35$
- (iii) $P(\text{At least one}) = 1 – \frac{20}{100} = 0.8$

Example:

In a survey:
• 40% of people read newspaper A
• 35% read newspaper B
• 15% read both newspapers

Find the probability that a person reads:

- (a) At least one newspaper
- (b) Neither newspaper

▼ **Answer/Explanation**

Step 1: $P(A) = 0.40, P(B) = 0.35, P(A \cap B) = 0.15.$

Step 2: At least one = $P(A \cup B) = P(A) + P(B) – P(A \cap B)$

= $0.40 + 0.35 – 0.15 = 0.60.$

Step 3: Neither = $1 – 0.60 = 0.40.$

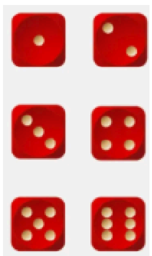
Mutually Exclusive Events

Mutually Exclusive Events

Mutually Exclusive Events

Two or more events are called **mutually exclusive** (or disjoint) if they cannot occur at the same time. In other words, the occurrence of one event means the other cannot happen in the same trial or experiment.

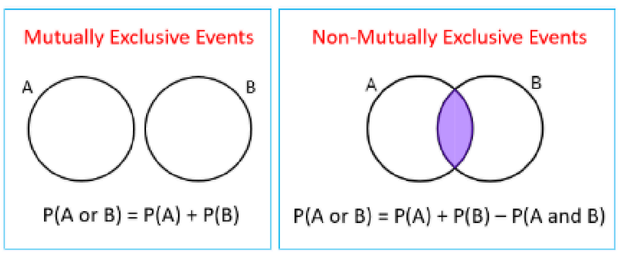
Example Meaning: If you roll a die, getting a "2" and a "5" on the same roll is impossible.



These outcomes are mutually exclusive.

Key Characteristics of Mutually Exclusive Events:

- No overlap in outcomes: $A \cap B = \emptyset$
- Cannot happen simultaneously
- Their probabilities are simply added: $P(A \cup B) = P(A) + P(B)$
- $P(A \cap B) = 0$



Comparison: Mutually Exclusive vs. Independent Events

Aspect	Mutually Exclusive	Independent(Non – Mutually Exclusive)
Occurrence	Cannot occur together	Can occur together
Formula	$P(A \cup B) = P(A) + P(B)$	$P(A \cap B) = P(A) \cdot P(B)$
Intersection	$P(A \cap B) = 0$	$P(A \cap B) \neq 0$

Real-Life Examples of Mutually Exclusive Events:

- Flipping a coin: Getting heads or tails (not both at once)
- Rolling a die: Getting a 2 or a 5 (you can only get one outcome)
- Selecting a student: A student cannot be both "present" and "absent" on the same day

Example:

Two events are defined as follows: A = rolling an even number on a die, B = rolling an odd number on a die. Are A and B mutually exclusive? Find $P(A \cup B)$.

▼ ▶Answer/Explanation

Step 1: A = {2, 4, 6}, B = {1, 3, 5}

Step 2: $A \cap B = \emptyset$, so they are mutually exclusive

Step 3: $P(A) = \frac{3}{6} = 0.5$, $P(B) = \frac{3}{6} = 0.5$

Step 4: $P(A \cup B) = P(A) + P(B) = 1$

Answer: Yes, mutually exclusive. $P(A \cup B) = 1$

Example:

A card is drawn from a standard deck. Let A = "drawing a heart", and B = "drawing a spade". Are A and B mutually exclusive? What is $P(A \cup B)$?

▼ ▶ Answer/Explanation

Step 1: A and B are different suits. No card can be both a heart and a spade.

Step 2: Mutually exclusive: $A \cap B = \emptyset$

Step 3: $P(A) = \frac{13}{52} = 0.25$, $P(B) = \frac{13}{52} = 0.25$

Step 4: $P(A \cup B) = 0.25 + 0.25 = 0.5$

Answer: Yes, they are mutually exclusive. $P(A \cup B) = 0.5$

Example:

In a class of 30 students: 12 like chess, 10 like football, 5 like both. Are the events "liking chess" and "liking football" mutually exclusive?

▼ ▶ Answer/Explanation

Step 1: Some students like both sports → overlap exists

Step 2: $A \cap B \neq \emptyset$

Answer: No, the events are **not mutually exclusive**.

Combined Events

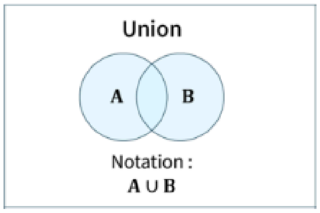
Combined Events

Combined Events

Combined events occur when two or more events happen together or in sequence in a single experiment. These events can be considered together to calculate probabilities.

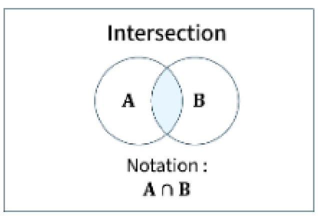
Types of Combined Events:

Union of Events (A or B): Event A happens OR Event B happens (or both).



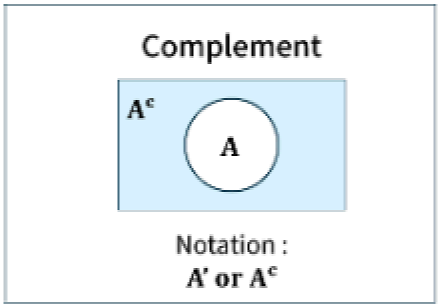
Formula: $P(A \cup B) = P(A) + P(B) – P(A \cap B)$

Intersection of Events (A and B): Both Event A and Event B occur.



Formula: $P(A \cap B)$

Complement of an Event: The event does NOT occur.



Formula: $P(A') = 1 – P(A)$

Key Notes:

- If A and B are **mutually exclusive**, then $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.
- If A and B are **independent**, then $P(A \cap B) = P(A) \times P(B)$.

Example:

A card is drawn from a standard deck. Find the probability that it is a **heart or a face card**.

▼ **Answer/Explanation**

Step 1: A = heart cards (13 cards), B = face cards (Jack, Queen, King in all suits = 12 cards)

Step 2: Intersection: cards that are both heart and face = 3 (J♥, Q♥, K♥)

Step 3: $P(A) = \frac{13}{52}, P(B) = \frac{12}{52}, P(A \cap B) = \frac{3}{52}$

Step 4: $P(A \cup B) = \frac{13}{52} + \frac{12}{52} – \frac{3}{52} = \frac{22}{52} = \frac{11}{26}$

Answer: $\frac{11}{26}$.

Example:

A die is rolled twice. Find the probability of getting **a 4 on the first roll and an even number on the second roll**.

▼ **Answer/Explanation**

Step 1: A = getting 4 on first roll = $\frac{1}{6}$

B = getting even number on second roll = $\frac{3}{6} = \frac{1}{2}$

Step 2: These are independent events, so $P(A \cap B) = P(A) \times P(B) = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$

Answer: $\frac{1}{12}$.

Example:

In a group of 50 students, 30 play football (F), 20 play basketball (B), and 10 play both. Find the probability that a student selected plays football or basketball.

▼ **Answer/Explanation**

Step 1: Use union formula: $P(F \cup B) = P(F) + P(B) – P(F \cap B)$

Step 2: $P(F) = \frac{30}{50} = 0.6,\ P(B) = \frac{20}{50} = 0.4,\ P(F \cap B) = \frac{10}{50} = 0.2$

Step 3: $P(F \cup B) = 0.6 + 0.4 – 0.2 = 0.8$

Answer: 0.8 or 80%.