Probability & Statistical Inference

Matthew Williams • Math • May 6, 2026

Probability measures uncertainty using numbers between 0 and 1. A probability near 0 means an event is unlikely, a probability near 1 means it is likely, and a probability of $\frac{1}{2}$ means the event has an even chance.

In CSEC, probability is often paired with data interpretation. You may calculate a theoretical probability from equally likely outcomes, compare it with experimental results, or make an inference from a sample. Always define the event clearly before counting favourable outcomes.

Probability measures the likelihood of an event occurring.

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

$$P(\text{event}) = \frac{n(E)}{n(S)}$$

Where:

- $n(E)$ = number of favorable outcomes
- $n(S)$ = total number of possible outcomes (sample space)
- **Range**: 0 dP d1 (or 0% to 100%)

Sample Space

The sample space is the full list of possible outcomes. If your sample space is incomplete, every probability based on it will be wrong.

The **sample space** is the set of ALL possible outcomes.

📋 **Example**

Rolling a dice:

Sample space = $\{1, 2, 3, 4, 5, 6\}$

$n(S) = 6$

Flipping a coin twice:

Sample space = $\{HH, HT, TH, TT\}$

$n(S) = 4$

[Mermaid:0]

Theoretical Probability

Theoretical probability is based on what should happen when outcomes are equally likely. It does not require an experiment.

Theoretical probability assumes all outcomes are equally likely (based on theory, not experiment).

📋 **Example**

P(rolling a 6 on a fair dice):

$$P(6) = \frac{1}{6}$$

P(getting a heads on a fair coin):

$$P(H) = \frac{1}{2}$$

P(drawing a red card from standard deck):

$$P(\text{red}) = \frac{26}{52} = \frac{1}{2}$$

Experimental Probability

Experimental probability is based on collected results. It may differ from theoretical probability, especially when the number of trials is small.

Experimental probability is based on actual experiments.

$$P(\text{event}) = \frac{\text{Number of times event occurred}}{\text{Total number of trials}}$$

📋 **Example**

Coin flip experiment: 100 flips, got heads 48 times:

$$P(\text{heads}) = \frac{48}{100} = 0.48$$

(Theoretical would be 0.5, experimental was 0.48,close!)

Complementary Events

The complement is the event not happening. It is often faster to calculate the complement and subtract from 1.

The **complement** of event E is "E does not happen."

$$P(E) + P(\text{not } E) = 1$$

$$P(\text{not } E) = 1 - P(E)$$

📋 **Example**

P(rolling a 6) = 1/6

P(not rolling a 6) = 1 - 1/6 = 5/6

Or directly: 5 ways to not roll a 6 out of 6 outcomes = 5/6

🧠 **Remember**

- **Sample space**: Set of ALL possible outcomes
- **Theoretical probability**: Based on logical reasoning
- **Experimental probability**: Based on actual data
- **P(E) + P(not E) = 1** (complementary events)

Interpreting and Making Inferences

Reading Data from Diagrams

Data diagrams must be read with attention to labels, scales, and units. A correct calculation can be wrong if the value was read from the wrong axis.

Statistical diagrams help us understand data without looking at all individual values.

📋 **Example**

From a pie chart showing favorite sports:

If soccer is 120° out of 360°:

$$\text{Proportion} = \frac{120^\circ}{360^\circ} = \frac{1}{3}$$

If 60 students total:

$$\text{Number who like soccer} = \frac{1}{3} \times 60 = 20$$

Making Inferences

An inference is a reasonable conclusion based on data, not a guess. It should mention what the data suggests and any limits of the sample.

Inference = drawing conclusions based on data.

📋 **Example**

Data: Average test score increases by 2 points per month over 6 months

Inference: "Study techniques are improving, or course content is better understood over time."

But be careful: Could be other reasons (easier tests, better teaching, student motivation, etc.)

Proportion or Percentage Above/Below a Value

Questions about above or below a value usually require counting a group first, then comparing it with the total.

📋 **Example**

From grouped data:

Class	Frequency
0-9	5
10-19	8
20-29	12
30-39	10
40-49	5

What proportion scored below 30?

Below 30 = 5 + 8 + 12 = 25
Total = 40

$$\text{Proportion} = \frac{25}{40} = \frac{5}{8} = 0.625 = 62.5\%$$

What proportion scored 20 or above?

20 or above = 12 + 10 + 5 = 27

$$\text{Proportion} = \frac{27}{40} = 0.675 = 67.5\%$$

Comparing Distributions

When comparing distributions, discuss both centre and spread. One dataset may have a higher average while another is more consistent.

Compare datasets using:

- Mean (central tendency)
- Range or IQR (spread)
- Shape (symmetric, skewed, bimodal)

Problem-Solving with Statistics

Statistics problem-solving usually combines calculation with interpretation. After finding the value, explain what it means in the situation.

Real problems require combining multiple skills.

📋 **Example**

Problem: A survey of 100 students' pocket money (in dollars):

[CodeBlock:0]

Questions:

- 1. **How many students have less than 40 dollars?**

8 + 15 + 25 = 48 students

- 2. **What is the median?**

Median position = 100 ÷ 2 = 50

Cumulative: 8, 23, 48, 78, ...

Median class is 40-49 (cumulative 78 includes position 50)

$$\text{Median} = 39.5 + \frac{50-48}{30} \times 10 = 39.5 + \frac{2}{30} \times 10 = 39.5 + 0.67 = 40.17$$

- 3. **What percentage earned 50 dollars or more?**

50 or more: 15 + 7 = 22

Percentage = (22 ÷ 100) × 100% = 22%

📝 **Exam Tip**

CSEC Statistics exam tips:

- 1. **Always identify class boundaries** for grouped data (not class limits)
- 2. **Label axes clearly** with title, units, and scale
- 3. **Use smooth curves** for ogives, not straight lines
- 4. **Show all working** for calculating mean, median, etc.
- 5. **Remember cumulative frequency uses upper boundaries**
- 6. **Read carefully**: "Below", "above", "at least", "more than" have different meanings
- 7. **For probability, find the sample space first**