

Cambridge AS & A Level · Psychology 9990 · Notes

Types of data and data analysis AS

Read online at pastopic.com/cambridge/psychology-9990/notes/types-of-data-and-analysis

Last checked 6 October 2026

Read first: [Research methods](#)

© 2026 Pastopic. Free for personal study. Please share the link, not the file.

What you need to know

By the end of this topic you should be able to:

1. **Types of data.** Describe **quantitative** and **qualitative** data, and **objective** and **subjective** data.
2. **Evaluate** each type of data as psychologists collect it, with examples from the core studies.
3. **Apply** this to a new study: suggest how to measure something so it gives the type of data you want, and say what is good or bad about that data in this study.
4. **Tables.** Present data in a table with labelled rows and columns, and read a table of results.
5. **Measures of central tendency.** Say what the term means; name, recognise and know how to find the **mode**, the **median** and the **mean**.
6. **Measures of spread.** Say what the term means; name, recognise and find the **range**; recognise, interpret and understand the **standard deviation**.
7. **Graphs.** Name, recognise, draw, change and interpret data from a **bar chart**, a **histogram** and a **scatter graph**.

The syllabus says you won't have to carry out calculations or understand statistical tests. You may still be asked to state a mode or a range from a short list, and to describe how a mean, median or range is found. Questions on histograms have also asked about the shape called the **normal distribution**, so it is included here.

Understand it

Quantitative and qualitative data

Quantitative data is **numerical**: a score, a count, a time, a rating on a scale, a percentage, or a total in each category. It tells you **how much** or **how many**.

Qualitative data is **descriptive**: **detailed, in-depth** information, such as what people say in their own words or a description of what they did. It gives **insight** into **why** people act as they do and **what it was like** for them.

Define qualitative data by its depth and detail. "Data in words" or "data that isn't numbers" is not accepted as a definition on its own.

The same topic can give either type, depending on how you collect it. "How nervous are you before a test, from 0 (not at all) to 10 (extremely)?" gives a number; "Describe how you feel before a test" gives a description.

	Quantitative	Qualitative
Where it comes from	Closed questions, rating scales, tallies of behavioural categories in an observation, test scores, physiological measures (heart rate, EEG, brain scans)	Open questions, transcripts of interviews, notes from an unstructured observation, what participants said
Strengths	Easy to compare groups and conditions and to analyse (averages, graphs); often objective , so it needs no interpretation	Gives detail and insight : people's own reasons, thoughts and feelings, including things the researcher didn't expect
Weaknesses	Can miss the reasons behind the numbers; closed questions limit what people can say	Hard to compare and summarise; has to be interpreted , so it can be subjective

A **number of words** counted in a description is quantitative: once you count, you have a number.

Objective and subjective data

Objective data is not affected by anyone's personal view: anyone who measured it would record the same thing. Times, counts of clearly defined behaviours, numbers of correct answers and heart rate are objective.

Subjective data depends on someone's **own point of view or interpretation**: the participant's (how they feel, what they remember) or the researcher's (deciding whether a behaviour "looked friendly", or how to score an answer).

Quantitative does **not** always mean objective. "Rate your happiness from 1 to 10" gives a number, but the number is the person's own judgement: one person's 6 may be another person's 8.

Subjectivity is a problem for **validity**, because the data may show someone's view rather than what really happened. Two cases come up often:

- **Recalling the past.** Memories are not fixed records. What happens afterwards can reshape a memory, details fade, and people may patch the gaps with invented details. An adult asked about early childhood might miss out real events entirely, or report events that never took place.
- **Researchers interpreting.** A researcher may read a different meaning into an answer from the one the participant intended. Two researchers scoring the same answers separately and comparing them (**inter-rater reliability**) shows whether this is a problem.

Core studies and their data

Core study	Data it collected
Milgram	Quantitative: the voltage each participant reached. Every participant went to at least 300 V , and 300 V was the most common point for those who refused to stop. Qualitative: what participants said while giving shocks (that they couldn't carry on, that it wasn't fair on the man, asking if he was all right) and their reactions, and what they said afterwards about taking part
Piliavin et al.	Quantitative: how many victims were helped (the cane victim 62 times, the drunk victim 19), how long help took, how many people were nearby. Qualitative: passengers' comments , for example saying they were too scared of the drunk man to help; there were more comments when the victim was drunk
Bandura et al.	Quantitative: observers counted each child's aggressive and non-aggressive responses; children were also rated for aggressiveness on five-point scales before testing; mean aggression scores were compared
Dement and Kleitman	Quantitative: EEG recordings, mean awakenings per night, the number of words in each dream narrative
Fagen et al.	Quantitative: time and number of cues given, and the percentage of success on each trained behaviour. Qualitative: an early observation that one elephant preferred salty water to plain water
Saavedra and Silverman	Qualitative: descriptions of the boy's fear of buttons
Pozzulo et al.	Quantitative: mean number of correct identifications per child
Baron-Cohen et al.	Quantitative: eyes test, AQ and IQ scores and ages, with means and standard deviations
Hassett et al.	Quantitative: mean time spent with, and how often each sex handled, plush and wheeled toys, shown on bar charts

Milgram shows the strength of each type. The voltages are objective and easy to compare: at a glance, everyone obeyed up to 300 V. The comments show what the numbers can't: how upset the participants were while they obeyed.

Tables

A results table needs a **heading for the rows** with its **categories**, and a **heading for the columns** with its **categories**. For a study timing how long two participants look at two kinds of picture:

Picture shown	Participant 1: time looking (s)	Participant 2: time looking (s)
Faces		

Picture shown	Participant 1: time looking (s)	Participant 2: time looking (s)
Landscapes		

Rows and columns can be swapped. Always say what is measured and its unit.

Measures of central tendency

A **measure of central tendency** is one number that sums up a set of scores by giving a **typical** or **average** score.

Take the number of items 9 people remembered from a list: 6, 9, 4, 9, 8, 5, 9, 7, 10.

- **Mode:** the score that occurs **most often**. Here 9 (three times). It is typical because it is the most common score.
- **Median:** put the scores **in order** and take the **middle** one: 4, 5, 6, 7, **8**, 9, 9, 9, 10, so 8. With an even number of scores, it is halfway between the middle two.
- **Mean:** **add up all the scores and divide by how many there are:** $\frac{67}{9} = 7.4$ (1 d.p.).

Measure	Use it for	Strength	Weakness
Mode	Categories (which subject, which word): the only one that works for them	Easy to find; shows the most common answer	Ignores most of the values; there may be several modes
Median	Rating scales , or scores with one or two extreme values	Not pulled up or down by an extreme score	Ignores the actual sizes of most scores
Mean	Scores on an equal scale (times, test marks where every item counts the same)	Uses the value of every score	Distorted by an extreme score

For example, times of 3, 4, 4, 5 and 34 seconds have a mean of 10 but a median of 4: one slow person drags the mean far above everybody else.

Measures of spread

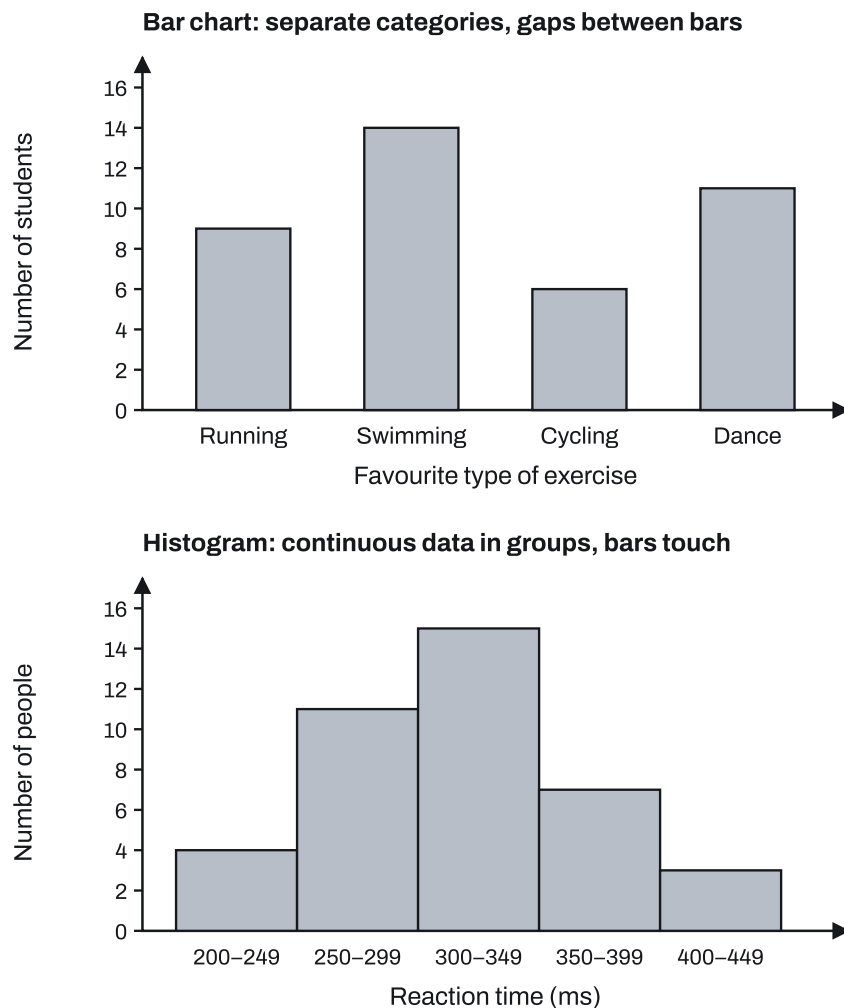
A **measure of spread** shows how much the scores **vary**: whether they are bunched together or far apart (dispersed).

- **Range:** the **highest score minus the lowest**. For the memory scores above, $10 - 4 = 6$. (Some mark schemes also accept adding 1.) It uses only the two extreme scores, so one unusual score makes it large.
- **Standard deviation (SD):** how far, **on average, the scores are from the mean**. A **small** SD means most scores are close to the mean; a **large** SD means they are spread out. You never calculate it.

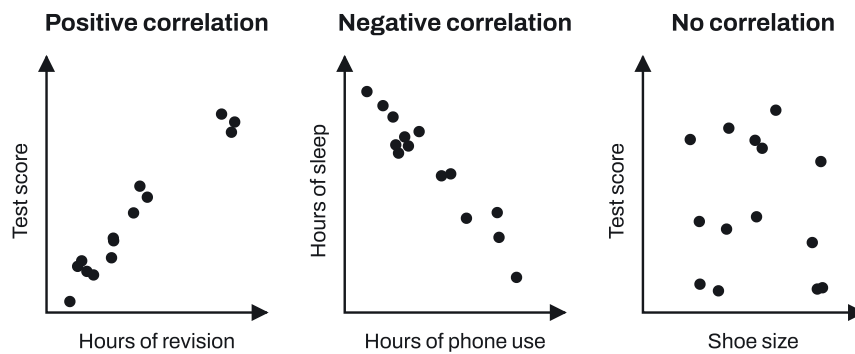
The SD tells you more than the range because it uses **every score**, not just the two end points.

In Pozzulo et al., every child was between 4 and 7 years old, but the adult sample covered a much wider span of ages, so the SD of the adults' ages was larger. In Baron-Cohen et al., Group 1 (adults with AS/HFA) had a higher mean age than Group 3 (students), 29.7 against 20.8, and a far larger SD, 14.5 against 0.8: the students were nearly all the same age, while Group 1's ages were spread out.

Graphs



- A **bar chart** shows data for separate **categories** (conditions, groups, types of thing). The categories go on the *x*-axis with **gaps** between the bars, because nothing lies between "swimming" and "cycling". Each bar's height is a number: a total, a percentage or a mean. It is the usual graph for **comparing conditions or groups**.
- A **histogram** shows how many scores fall in each **interval of a continuous variable** (time, age, temperature). The intervals go on the *x*-axis in order and the bars **touch**, because one interval runs straight into the next. The *y*-axis is the **frequency**: the number of participants.
- A **scatter graph** shows a **correlation**: one co-variable on each axis, and **one point per participant**.

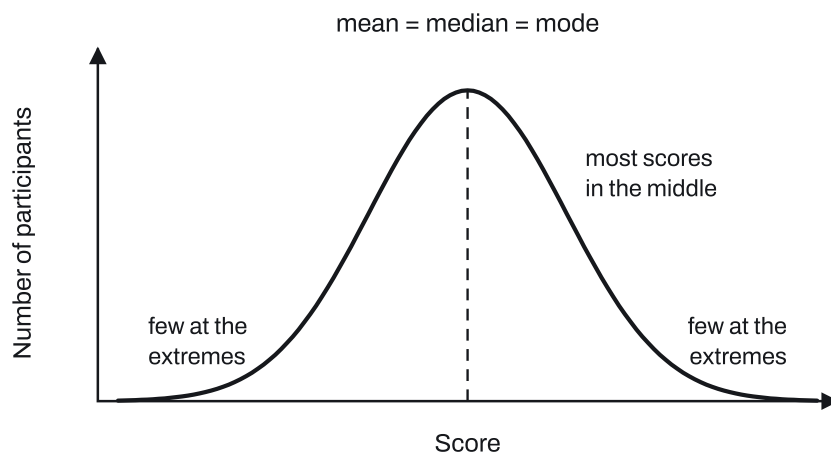


- **Positive:** as one variable increases, so does the other (points rise to the right).
- **Negative:** as one increases, the other decreases (points fall to the right).
- **No correlation:** no pattern.

The **strength** is how closely the points lie to a straight line. A correlation never shows that one variable **causes** the other: a third variable could affect both, or the cause could run either way. Only data on a **scale** (scores, times, ratings) can be correlated; categories such as "tidy / untidy" or "yes / no" can't.

The normal distribution

Many human characteristics measured in a large group (reaction time, scores on a well-designed test) give a histogram with a typical shape: the **normal distribution**.



- It is **bell-shaped** and **symmetrical**, with half the scores either side of the middle.
- The **mean, median and mode are equal**, at the highest point.
- **Most scores are in the middle**, with fewer and fewer towards the extremes (the tails).

Baron-Cohen et al. graphed the percentage of participants getting each eyes test score, and the shape was a normal distribution.

Key facts and formulas

There is no formula list for 9990, and you won't be asked to calculate. "Learn it" means you must remember it.

Fact	Status
Quantitative data: numerical (counts, totals in categories, times, scores, ratings), easy to compare and analyse	Learn it
Qualitative data: descriptive, detailed, in-depth data giving insight into reasons and feelings	Learn it
Objective data: not affected by personal views, so anyone would record the same; subjective data: depends on a person's own point of view or interpretation	Learn it
Measure of central tendency: one typical or average score that sums up a set of data	Learn it
Mode: the most frequent score; median: the middle score when the scores are in order; mean: the total of the scores divided by the number of scores	Learn it
$\text{mean} = \frac{\text{total of all the scores}}{\text{number of scores}}$	Learn it
Mode for categories; median for rating scales or data with extreme scores; mean for scores on an equal scale	Learn it
Measure of spread: how much the scores in a set vary	Learn it
$\text{Range} = \text{highest score} - \text{lowest score}$	Learn it
Standard deviation: the average difference of the scores from the mean; larger SD = scores more spread out	Learn it
Bar chart: separate categories, bars with gaps; histogram: continuous data in intervals, bars touching, frequency on the y -axis; scatter graph: a correlation, one point per participant	Learn it
Normal distribution: bell-shaped, symmetrical, mean = median = mode at the highest point	Learn it

How to do it

We have tagged 94 Paper 2 questions on this topic, from 2021 to March 2026. 8 of them are mainly about other topics (sampling, designs, hypotheses) and aren't counted below. Of the other 86, 38 have parts of more than one type, so the counts add up to more than 86.

Question type	Questions
Evaluate a measure or way of collecting data (validity, reliability, practical problems)	20

Question type	Questions
Suggest a measure / operationalise a variable (quantitative score, rating scale, behavioural categories)	16
Draw a conclusion from a graph or table	10
Calculate, or describe how to calculate, the mean, median, mode or range	10
Describe the data a core study collected (qualitative or quantitative)	9
Plot a graph (bar chart or histogram) from given data	8
Define or interpret measures of spread (range, standard deviation)	8
Define or identify quantitative vs qualitative data	7
Strengths and weaknesses of quantitative or qualitative data	7
Choose or justify a measure of central tendency	7
Name or justify the right type of graph	7
Sketch a correlation on labelled axes	6
Interpret a correlation (direction, strength, suitable data, not causation)	4
Label or draw axes for a graph (no data)	4
Normal distribution (sketch, name, features)	4
Subjectivity and objectivity	4
Write or identify open and closed questions	4
Draw a table to record data	4
6-mark "Describe ... using any example(s)"	2

1. Evaluate a measure or way of collecting data (20 questions)

1. **Name the problem or strength** (1).
2. **Link it to the scenario**: say what it does to the data in this study, often "so the data would ..." (1). A point that fits any study usually caps at 1.
3. If asked, say whether it is a problem of **validity** (not measuring what it should) or **reliability** (not consistent).

Measure	Valid because...	Not valid because...
Questionnaire (paper or online)	Private, so people are more honest than face to face (less social desirability)	People may misunderstand a question and nobody is there to explain

Measure	Valid because...	Not valid because...
Interview	If a question is unclear, the participant can ask and have it clarified on the spot	Face to face, people may give answers that make them look good
Closed question / rating scale	Numbers are objective and easy to compare between groups	Fixed choices may not include the answer the person wants to give; different people use the numbers differently
Open question	People can explain their own reasons in detail	Turning a free answer into a number is a judgement call, so two researchers may score it differently (subjective)
Self-report of behaviour at home (sleep, exercise)	Real-life setting	People may misremember or lie; the home setting (noise, pets, children) can't be controlled
Observation	Records what people do, not what they say	Feelings can't be seen; a behaviour or emotion may be misread
A two-option judgement (tidy / not tidy)	Quick and simple	Can't tell different levels apart

"It isn't generalisable" is about the sample, not the measure, and earns nothing here.

Variations you will meet: a valid **and** a not-valid point about the same method; comparing questionnaires with interviews or two kinds of brain scan; whether a measure is fair to two species of different sizes (if the equipment is physically easier for the larger species, its scores reflect body size as well as learning).

2. Suggest a measure / operationalise a variable (16 questions)

1. **Make it concrete.** Say exactly what is counted, timed or rated, and how:

- a **count** of a clearly defined behaviour (times a child looks at the screen);
- a **time** (seconds until a puzzle is solved, minutes with each object);
- a **test score** (items correctly recalled out of 20);
- a **rating scale** with both ends labelled ("from 0 = wide awake to 10 = falling asleep");
- **behavioural categories** for an observation (smiling, frowning, turning away);
- a **physiological** measure (heart rate, EEG).

2. **Say what the number means** for the variable: "the fewer trials it needs, the more easily it has learned". This often earns the second mark.

3. **Read the restrictions:** some questions forbid self-report or interviews.

Quantitative data from a questionnaire or interview: either a **closed question with a scale** ("On a scale of 0 to 10, how much would you like to join a new club?"), or an **open question whose answer the researcher then scores** on a scale.

3. Draw a conclusion from a graph or table (10 questions)

1. **Find the pattern:** which condition or group is higher, and where.
2. **Write a general statement about people that compares:** "people remember more words in a quiet room than in a noisy one". 1 mark for any conclusion, 2 for a full comparison across both variables.
3. **Don't just restate numbers.** "Group A scored 12 and group B scored 8" is a result, not a conclusion, and scores 0.

Variation: explain why someone's conclusion is wrong: check every bar, since a difference may appear in only one condition.

4. Calculate, or describe how to calculate, the mean, median, mode or range (10 questions)

- **"State the mode":** the score that appears most often. Don't give the median or mean.
- **"Describe how the mean would be calculated in this study":** use the study's own words: "add up all of [each student's] [weekly test scores] and divide by the number of [tests]" (1 for the method, 1 for the link). Dividing by the wrong number (children rather than possible identifications, say) loses the link mark.
- **Median:** put the [times] in order and take the middle one.
- **Range:** subtract the lowest [hours asleep] from the highest.

5. Describe the data a core study collected (9 questions)

1. **Identify** what was recorded (1): Milgram's voltages or comments; Piliavin et al.'s helping counts or passengers' comments; Bandura et al.'s counts of aggressive acts by observers.
2. **Add detail or an example** (1): an example comment (it needn't be word for word), or how the data was used ("more comments when the victim was drunk").

Learn the table in "Understand it". For Milgram, 300 V was the highest voltage that every participant reached.

6. Plot a graph (bar chart or histogram) from given data (8 questions)

1. **Choose the graph.** Categories or conditions → bar chart with gaps. Intervals of a continuous variable (age groups, temperature bands) → histogram with bars **touching**; gaps can cost a mark.
2. **Label both axes** with the variable and unit: "Age group (years)", "Mean score out of 20" (1 each).
3. **Write the categories or scale:** each category under its bar; numbers up the *y*-axis at even steps, at least as high as the largest value (1 each).
4. **Plot the heights accurately**, and plot only what is asked.

7. Define or interpret measures of spread (8 questions)

1. **Define:** SD is a measure of spread (1) based on the difference of scores from the mean (1). A measure of spread shows how much scores vary; "how spread out the scores are" repeats the question.

2. **Read a table:** the larger SD means the scores are **more spread out**. "The SD was bigger" or "they deviated more" scores 0. A mean says which group scored higher; an SD says which group was more varied.
3. **Explain why one SD is larger**, linked to who was in each group: one chosen from a narrow band (young children, students), the other from a wide one (adults of all ages).
4. **SD compared with range:** SD uses every score, not just the two extremes, so one odd score distorts it less.

8. Define or identify quantitative vs qualitative data (7 questions)

1. **Define** each: quantitative = numerical (a count, totals in categories); qualitative = descriptive, detailed, in depth.
2. **Identify** a given example and **justify**: "quantitative, because it is a number of words".
3. "Explain the difference" needs **both** definitions. Saying one is objective and the other subjective doesn't count as the difference.

9. Strengths and weaknesses of quantitative or qualitative data (7 questions)

1. **Strength or weakness** (1): quantitative is easy to compare and analyse, and objective; qualitative gives detail and shows thoughts and feelings.
2. **Link to the study** (1): "the voltages made it easy to see that everyone obeyed to 300 V"; "the comments showed how upset the participants were".

10. Choose or justify a measure of central tendency (7 questions)

1. **Look at the data:** categories → mode; a rating scale → median (the mean is often accepted); scores on an equal scale → mean.
2. **Give a reason linked to the data:** "the answers are categories, so you can only find which is most common"; "every question is equally difficult, so the mean, which uses every score, is best".
3. **Why it is a measure of central tendency:** it gives one **typical** score for the whole set; for the mode, the most frequent score.
4. **Mean better than mode:** it uses the value of every score, while the mode only uses how often each occurs. **Mode instead of mean:** the data is in categories or totals, or there are extreme scores.

11. Name or justify the right type of graph (7 questions)

1. **Name it:** usually "bar chart".
2. **Justify:** the x -axis variable is separate **categories** (male and female, two toy types, two conditions), so bars side by side let you compare them; say what goes on each axis for a further mark.

A histogram for categories is wrong.

12. Sketch a correlation on labelled axes (6 questions)

1. **Label both axes** with the two co-variables, either way round (1). For 3–4 marks add units or a scale ("hours" on a sleep axis, words or numbers on a health axis).
2. **Plot points or a line** going the right way: up to the right for positive, down to the right for negative (1).

13. Interpret a correlation (4 questions)

- **Direction:** both rise together = positive; say it for these variables.
- **Strength:** how close the points lie to a line.
- **Unsuitable data:** a yes/no or tidy/untidy measure gives categories, which can't be correlated; it needs a scale.
- **Not causation:** a correlation can't show which variable affects the other; a third variable could cause both. Give a linked example.

14. Label or draw axes for a graph (4 questions)

x = the IV or the groups (Monday / Friday, species 1 / species 2), with its categories; y = the DV with units ("Number of people who ...", "Time to learn (seconds)"), with numbers. One mark each for x label, x categories, y label and y numbers.

15. Normal distribution (4 questions)

- **Features:** any two of bell-shaped; symmetrical; mean, median and mode equal at the highest point; most scores in the middle and few at the extremes.
- **Sketch:** x -axis "Score on [the scale]" with its range (for example 0–5); y -axis "Frequency" or "Number of participants"; a symmetrical bell.
- **Name it** from a graph: "normal distribution".

16. Subjectivity and objectivity (4 questions)

1. **Define** in your own words: subjectivity = a judgement shaped by the person making it (their feelings, expectations or memories), so someone else could judge the same thing differently. "Not being objective" scores 0.
2. **Apply** it (1 + 1 detail): recalled events may be distorted by time or later events; a researcher may interpret an answer differently from how it was meant.
3. **Objective alternative:** a measure nobody has to interpret (a count, a time, heart rate).

17. Write or identify open and closed questions (4 questions)

- **Closed:** give the answer options or the scale with its end points ("from 1 = not at all to 5 = very"). Gives quantitative data.

- **Open:** ask for a detailed answer with "describe", "explain" or "why". Gives qualitative data.
- Write each about the scenario's topic.

18. Draw a table to record data (4 questions)

One mark each: **row heading**, **row categories**, **column heading**, **column categories**. Use the study's own words ("Type of object" with each object listed; "Time spent with object (seconds)" over Participant 1 and Participant 2). Rows and columns can be swapped. To turn a graph into a table, use its axis labels as headings.

19. 6-mark "Describe ... using any example(s)" (2 questions)

Asked on measures of central tendency and on quantitative and qualitative data.

1. **Up to 2 marks per term** for what it is and a detail (how each average is found; where each type of data comes from).
2. **Up to 4 marks for examples**, at most 2 per term, from core studies or your own. Without examples, or covering only one term, the most you can get is 4.
3. Examples: Bandura et al.'s or Pozzulo et al.'s means; Milgram's voltages and comments; Saavedra and Silverman's descriptions of the boy's phobia. There are no marks for evaluation.

Worked examples

Example 1

Tomas reads 9 people a list of 20 words and asks them to write down as many as they can. Their scores are:

13, 16, 10, 16, 12, 15, 16, 9, 11

- (a) State the mode for these scores. **[1]**
 - (b) Describe how Tomas would find the median for these scores. **[2]**
 - (c) Explain why the mean would be a suitable measure of central tendency for Tomas's data. **[2]**
 - (d) State the range of the scores. **[1]**
- (a) 16 (1).** It occurs three times; no other score occurs more than once.
- (b) Put the 9 scores in order from lowest to highest (1):** 9, 10, 11, 12, 13, 15, 16, 16, 16, then take the middle (5th) score, which is 13 **(1)**.
- (c) The scores are on an equal scale**, because every word recalled counts as one **(1)**, so the mean can use the value of every score, adding all 9 and dividing by 9, which makes it more sensitive than the mode **(1)**. (The mean is $\frac{118}{9} = 13.1$ to 1 d.p.; you won't be asked to work it out.)

(d) $16 - 9 = 7$ (1).

Example 2

Keisha measured the reaction times of 40 people.

Reaction time (ms)	200–249	250–299	300–349	350–399	400–449
Number of people	4	11	15	7	3

(a) Name the most suitable type of graph for Keisha's data and explain why. [2]

(b) Plot the graph. You must label the axes. [4]

(c) State the interval in which the mode lies. [1]

(a) A histogram (1), because reaction time is continuous: the intervals follow on from each other, so the bars should touch (1).

(b) The bottom graph in the first diagram. x -axis labelled "Reaction time (ms)" (1) with the five intervals in order (1); y -axis labelled "Number of people" (1); bars at the right heights, touching (1).

(c) 300–349 ms (1), the interval with the most people (15).

Example 3

Ayesha predicts a negative correlation between the minutes students spend outdoors each day and their stress score (a rating from 0 = no stress to 10 = extremely stressed).

(a) Sketch a scatter graph to show the result Ayesha predicts. You must label the axes. [2]

(b) Ayesha finds a strong negative correlation and concludes that spending time outdoors reduces stress. Explain why she cannot conclude this. [2]

(c) Explain whether Ayesha's stress score is objective or subjective data. [2]

(a) Axes labelled "Minutes outdoors per day" and "Stress score (0–10)", either way round (1); points falling from top left to bottom right (1).

(b) A correlation only shows that two variables are related, not that one causes the other (1). A third variable could affect both: students with fewer exams may go outside more and also feel less stressed; or stressed students may stay in to work (1).

(c) Subjective (1): each student rates their own stress from their own point of view, so one student's 7 may feel the same as another's 4, even though it is a number (1).

Example 4

Dr Rossi timed 40 people solving the same puzzle. Half were given a printed hint card first (Condition A); the other half had no hint (Condition B).

	Mean time (seconds)	Standard deviation
Condition A	42.5	3.1
Condition B	47.9	12.6

- (a) Compare the average solving times in the two conditions. **[1]**
- (b) Use the standard deviations to compare how consistent the times were in each condition. **[1]**
- (c) Suggest one reason why the times in Condition B varied more than those in Condition A. **[2]**
- (d) Explain one advantage of using the standard deviation rather than the range. **[2]**
- (a)** People given the hint card solved the puzzle faster on average, 42.5 s against 47.9 s **(1)**.
- (b)** Condition A's times were tightly bunched around their mean, while Condition B's were far more spread out **(1)**.
- (c)** With the hint, everyone followed the same route to the answer, so their times were alike **(1)**. Without it, some people spotted the trick straight away while others tried several wrong approaches first, so times ran from quick to very slow **(1)**.
- (d)** The standard deviation uses every score, while the range uses only the highest and lowest **(1)**, so one person who took a very long time would stretch the range but affect the standard deviation less **(1)**.

Common mistakes

- **Defining qualitative data as "words" or "not numbers"**. Mark schemes want descriptive, detailed or in-depth data. Saying "quantitative is objective, qualitative is subjective" doesn't earn the difference either.
- **Thin descriptions of qualitative data**. In March 2025, answers often only said that depth comes from open questions; add other sources and an example.
- **Giving the median or mean when asked for the mode** (March 2026), and repeating a term instead of explaining it.
- **The wrong measure for test scores**. For a test where every question is equally hard, the answer is the mean; "calculate the average", standard deviation or a histogram is not a measure of central tendency (March 2025).
- **A histogram for categories**. Category data needs a bar chart (March 2025, June 2023).
- **IV and DV on the wrong axes** (March 2023): the IV or groups go along the bottom, the DV up the side.
- **Restating the numbers when asked for a conclusion**. A result is the data; a conclusion interprets it (March 2021, 2023 and 2025).

- **Saying a correlation shows cause**, or trying to correlate data that isn't on a scale (March 2023).
- **Normal distribution with the wrong *y*-axis**. It must be frequency or the number of participants, not the score again (June 2022).
- **"The SD is bigger" or "the data deviated more"** without saying the scores were more spread out.
- **Saying a bar chart collects or analyses data**. A graph only displays data (March 2021, March 2026).
- **Thinking every number is objective**. A rating is quantitative but can be subjective.

How to get the marks

- **"State", "name", "identify"**: one word or number, and it must be the exact answer ("bar chart", "6", "quantitative").
- **"Outline what is meant by..."**: a definition that doesn't reuse the term.
- **"Explain..."**: the point (1), then why, or what it means in this study (1).
- **"Suggest"**: apply your answer to the scenario; a generic suggestion usually earns only 1.
- **"Describe how the mean / median / range would be calculated in this study"**: the steps, using the study's own variable.
- **"You must label the axes"**: a label on each axis, plus categories or a scale, or the marks for them are lost.
- **6-mark "Describe..., using any example(s)"**: both terms, at least one example, no evaluation.

Quick check

1. A gym manager wants to know what members think of a new class timetable. Suggest one way to collect quantitative data and one way to collect qualitative data, and give one strength of each type in this study.
2. A researcher counts how many times 7 toddlers point at a picture book: 5, 8, 8, 4, 10, 8, 6. State the mode and the range.
3. Name the most suitable graph for: (a) the number of people choosing each of four breakfast cereals; (b) the heights of 60 students in 5 cm intervals; (c) the relationship between hours of practice and goals scored.
4. People rate how anxious they feel before a driving test from 0 to 10. Explain why this is subjective data, and suggest an objective measure of anxiety.
5. Hassan times how long 200 adults take to read one page, and his histogram shows a normal distribution. Describe its shape, and say where the mean, median and mode are.

Answers

1. Quantitative: "Rate the new timetable from 1 = much worse to 5 = much better". Qualitative: "Describe what you like or dislike about the new timetable". Quantitative strength: the ratings can be compared, for example morning against evening members. Qualitative strength: members give their own reasons (a class clashes with work), showing the manager what to change.
2. Mode = 8 (it occurs three times). Range = $10 - 4 = 6$.
3. (a) Bar chart: the cereals are categories. (b) Histogram: height is continuous, so the bars touch. (c) Scatter graph: it shows a correlation between two co-variables.
4. Each person rates their own anxiety from their own point of view, so the same number can mean different things to different people. Objective: heart rate measured with a monitor just before the test.
5. Bell-shaped and symmetrical: the tallest bars are in the middle, so most adults took a middling time, and the bars get shorter towards both ends, so few were very fast or very slow. The mean, median and mode are equal, under the tallest bar.

Past questions to try

- [9990/22/O/N/23 Q9](#): the right measure of central tendency and graph for two questions, and sketching a normal distribution.
- [9990/21/M/J/24 Q7](#): mode or mean for two survey questions, how to find the range, and why the standard deviation tells you more.
- [9990/23/M/J/25 Q3](#): sketch a positive correlation with labelled axes, and problems collecting sleep data at home.
- [9990/23/M/J/25 Q2](#): what subjectivity means, and why it is a problem when people recall past events.