Statistics & Probability · Unit S01 · Foundations

Data All Around

Every number in every chart you will ever meet started life as somebody writing something down. This unit is about that moment: what data is, where it comes from, which kind you are holding, and how to get a messy list of twenty answers into a table that can be read at a glance.

Foundations (MYP 1–3) · about 45 minutes · 60 practice questions · Criteria A, B, C and D

BY THE END OF THIS UNIT YOU CAN

• **Describe** what data is, and say whether a set of data is **primary** (you collected it) or **secondary** (somebody else did)

• **Classify** a variable as **qualitative** (words) or **quantitative** (numbers)

• **Classify** a quantitative variable as **discrete** (counted) or **continuous** (measured), and give the reason

• Turn a raw list into a **tally chart**, bundling in fives, and then into a **frequency table**

• **Complete** a frequency table and check it against its total

• Build and read a **two-way table**, and tell a cell from a row total

• **Read** a completed table back: how many of one kind, how many altogether, how many more than another

• **Draw** a first bar chart from a frequency table, with the frequency axis starting at 0

THE ONE IDEA BEHIND ALL OF IT

Data is just answers, written down. Everything else in statistics — every average, every chart, every probability — is something you do to answers that somebody wrote down. So the first skill is not calculating. It is **classifying**: knowing what kind of answer you are holding, because the kind decides the table you build and the chart you may draw.

collect → classify → count → chart

and the frequencies must always add back to the number you started with

WHERE THIS UNIT SITS

This is the first unit of the Statistics & Probability strand. It needs no new number work at all — counting and adding whole numbers is the whole of the arithmetic. Number Unit N01 *Know Your Numbers* covers that, and nothing from it is retaught here.

What comes next. S02 asks how to word the question you collect the data with, S03 asks whether the data can be trusted, and S04 draws the full family of charts — pie, pictogram, stem-and-leaf — from tables exactly like the ones you build here.

Where the numbers in this unit come from. The traffic count is a list of twenty vehicles, printed in full; the two-way table is sixteen named-nothing learners; the sibling table is twenty-four. Every count in every chart on this page can be checked by hand against the list it came from, and that is deliberate.

CONCEPT 1 · WHAT COUNTS AS DATA

primary and secondary

Data is a set of facts you have collected, written down so they can be counted. A **variable** is the thing you are collecting — way to school, shoe size, favourite colour. The individual answers are the **data values**. If you collected them yourself, the data is **primary**. If you took them from somebody else — a textbook, a website, last year's class — it is **secondary**. Primary data you can vouch for; secondary data you have to trust.

> **EXAMPLE.** You and your class count vehicles at the gate for twenty minutes. That is **primary** data. You then look up how many combis are licensed in Gaborone in a reference book. That is **secondary** data. Both are useful; only the first one is yours.

CONCEPT 2 · WORDS OR NUMBERS?

qualitative and quantitative

A variable is **qualitative** when its answers are words or categories — combi, Setswana, blue. It is **quantitative** when its answers are numbers you can do arithmetic with — 7 vehicles, 152 cm, 4 siblings. The test is not whether digits appear. A jersey number is written with digits, but adding two jersey numbers gives nothing, so it behaves as a category.

Worked example 1 · Sort the school bag

1	Shoe size — counted, measured or neither?	**quantitative, discrete**
2	Favourite colour — counted, measured or neither?	**qualitative**
3	Height — counted, measured or neither?	**quantitative, continuous**
4	Number of brothers and sisters — counted, measured or neither?	**quantitative, discrete**
5	Home language — counted, measured or neither?	**qualitative**
6	Mass of the school bag — counted, measured or neither?	**quantitative, continuous**
7	Number of pencils in the bag — counted, measured or neither?	**quantitative, discrete**
8	Time taken to get to school — counted, measured or neither?	**quantitative, continuous**

Eight variables, three questions each time: words or numbers, and if numbers, counted or measured. Nothing else is needed, and nothing else counts as an answer.

EXAMPLE. “How did you get to school?” gives combi, walk, car — **qualitative**. “How many minutes did it take?” gives 12, 25, 8 — **quantitative**. Same learners, same morning, two different kinds of data.

CONCEPT 3 · COUNTED OR MEASURED?

discrete and continuous

Quantitative data splits again. **Discrete** data is **counted**: it lands on separate, fixed values with nothing in between — 3 goals, 4 siblings, size 6 shoes. **Continuous** data is **measured**: between any two values there is always another one, and measuring more carefully just adds decimal places — height, mass, time. The question to ask is never “is it a number?” but **“did I count it or measure it?”**

Worked example 5 · Spot the impostor

1	Number of learners in a class — counted or measured?	**counted, so discrete**
2	Number of goals scored — counted or measured?	**counted, so discrete**
3	Length of a pencil — counted or measured?	**measured, so continuous**
4	Number of pages in a book — counted or measured?	**counted, so discrete**
5	So which of the four was wrongly called discrete?	**length of a pencil**
6	Give the reason.	**a pencil can be any length; measuring more carefully just adds decimal places**

Three of the four are counts of separate things. The impostor is the one you would reach for a ruler to find, and reaching for a ruler is the test.

EXAMPLE. Shoe size looks like a measurement, and it is not. Sizes go 5, 6, 7 — and half sizes if the shop stocks them — but never 6.31. It is **discrete**. The length of the foot inside the shoe is **continuous**.

CONCEPT 4 · TALLY IT, TABLE IT

from a raw list to a frequency table

A raw list is unreadable. A **tally chart** makes it countable: one mark per value, and every fifth mark drawn across the previous four so the marks read in fives. Counting in fives is the whole point — it is why the bundles exist. A **frequency table** then records the totals, one row per category, with the **frequency** being how many times that category occurred. The frequencies must add up to the number of values you started with, and checking that is not optional.

EXAMPLE · the raw list, in full. Twenty vehicles went past the gate of a school, in this order: Car; Car; Combi; Car; Bakkie; Car; Combi; Motorbike; Combi; Combi; Combi; Bakkie; Combi; Bakkie; Combi; Lorry; Motorbike; Car; Motorbike; Bakkie.

Tally chart · vehicles past the gate in twenty minutes

Category	Tally	Frequency
Combi	𝍸 ||	7
Car	𝍸	5
Bakkie	||||	4
Lorry	|	1
Motorbike	|||	3
Total		20

Twenty vehicles, one tally mark each, bundled in fives.

One mark per vehicle, bundled in fives. The tally column and the frequency column say the same thing twice, which is the check.

Frequency table · the same twenty vehicles

Vehicle type	Frequency
Combi	7
Car	5
Bakkie	4
Lorry	1
Motorbike	3
Total	20

The finished frequency table. The five frequencies add to 20, the number of vehicles in the raw list above.

Complete the table

Vehicle type	Frequency
Combi	7
Car	5
Bakkie	
Lorry	1
Motorbike	
Total	20

Two frequencies have been rubbed out. The raw list is above, so count them again.

Two frequencies rubbed out. Count them again from the raw list — not from the total, which would only ever give you the pair.

Worked example 2 · Tally the traffic

1	Write the five vehicle types down the side of a tally chart.	**five rows**
2	Work along the list of 20, one mark per vehicle, bundling every fifth mark across the four before it.	**20 marks in all**
3	Count the Combi row in fives.	**Combi: 7**
4	Count the Car row.	**Car: 5**
5	Count the Bakkie, Lorry and Motorbike rows.	**4, 1 and 3**
6	Add the five frequencies and check against the list length.	**7 + 5 + 4 + 1 + 3 = 20**

The check at the end is the whole discipline. If the frequencies had come to 19 or 21, the tally would be wrong and the chart drawn from it would be wrong too.

EXAMPLE. Twenty vehicles go past. You do not write “combi” twenty times; you make one mark per vehicle in the right row and bundle every fifth mark across the other four. At the end you count in fives, and the five row totals must come to 20.

CONCEPT 5 · TWO WAYS AT ONCE

the two-way table

A **two-way table** classifies one set of data by two variables at once. Every learner is counted exactly once, in exactly one cell. The **row totals** tell you about the first variable, the **column totals** about the second, and the **grand total** sits in the corner where the two agree. If the row totals and the column totals do not give the same grand total, something has been counted twice or missed.

Two-way table · gender and way to school

	Walk	Combi	Car	Total
Girls	3	4	2	9
Boys	2	3	2	7
Total	5	7	4	16

Sixteen learners of Form 1C, classified two ways at once.

Every learner sits in exactly one cell, so the grand total can be reached down the rows and across the columns.

Worked example 3 · Two ways at once

1	Sixteen learners of Form 1C. Put gender down the side and way to school across the top.	**a 2 by 3 grid**
2	Count the girls who walk.	**Girls, Walk: 3**
3	Count the girls who come by combi, and the girls who come by car.	**4 and 2**
4	Add the girls' row.	**3 + 4 + 2 = 9**
5	Add the boys' row.	**7 boys**
6	Add the two row totals, then add the three column totals, and check they agree.	**9 + 7 = 16, and 5 + 7 + 4 = 16**

Every learner is in exactly one cell, so the grand total can be reached two ways. When the two ways disagree, the table is wrong — and that is the

only warning you get.

EXAMPLE. Sixteen learners, classified by gender and by way to school. The cell where *Girls* meets *Combi* holds 4. That is four girls who came by combi — not four girls altogether, and not four combi passengers altogether.

CONCEPT 6 · READ IT BACK

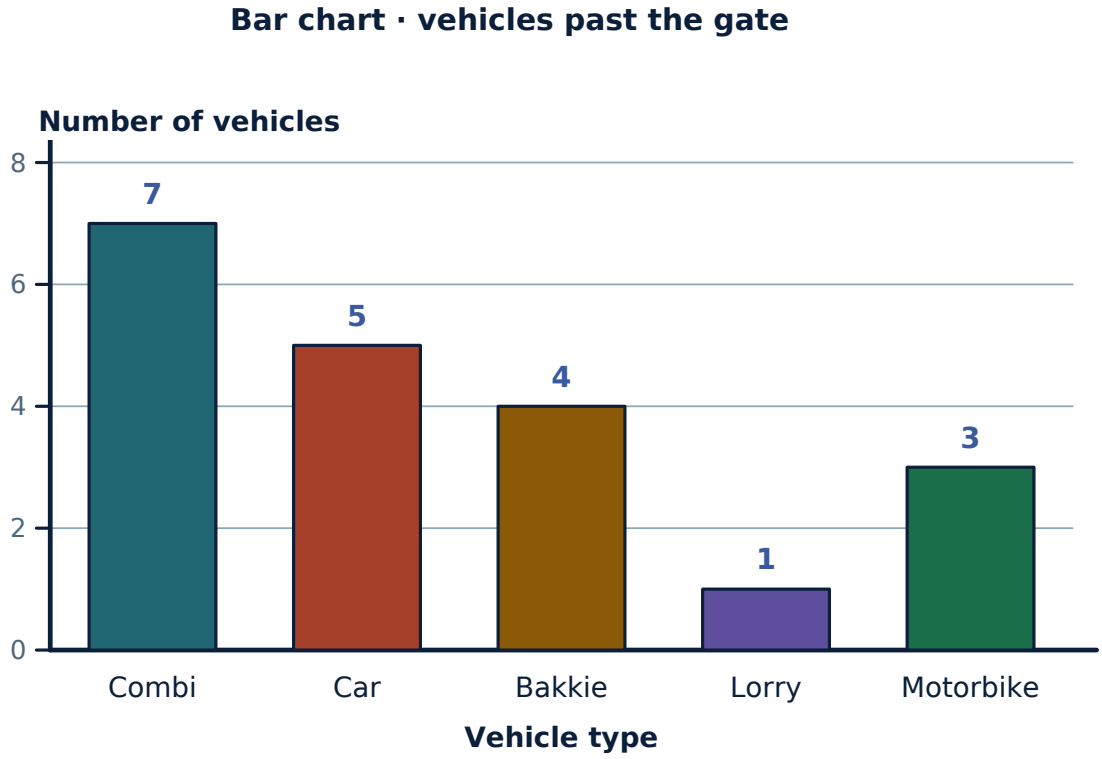
answering questions from a table

Reading a table back is a skill in itself, and the questions come in four shapes: **how many of one kind** (read one cell), **how many altogether** (add the frequency column), **how many more** (subtract two frequencies), and **how many are not** (total minus one frequency). A **bar chart** is the same table drawn: one bar per category, the frequency axis starting at 0, and equal gaps between the bars because the categories are separate things.

A completed frequency table · 24 learners

Brothers and sisters	Number of learners
0	3
1	7
2	6
3	5
4	2
5	1
Total	24

A completed frequency table, ready to be read back.



The same table drawn. One bar per category, equal gaps, and the frequency axis starting at 0.

Worked example 4 · Read it back

1	How many learners have exactly 2 brothers and sisters? Read the row.	**6 learners**
2	How many learners answered altogether? Add the frequency column.	**24 learners**
3	How many have more than 3? Add the rows for 4 and 5.	**2 + 1 = 3**
4	How many do NOT have exactly 1? Total minus that frequency.	**24 − 7 = 17**

Four question shapes, four different moves: read a row, add the column, add two rows, subtract from the total. Deciding which move the question wants is most of the mark.

EXAMPLE. From the sibling table: **6** learners have exactly 2 brothers and sisters; **24** learners answered altogether; **1** more learner has 1 than has 2; and **18** learners do not have exactly 2, because 24 − 6 = 18. Every one of those four answers is a different move on the same table.

THE CLASSIFYING ROUTINE

1. Who collected it? You, or somebody else? That is primary or secondary, and it is the first thing to say about any dataset.

2. Words or numbers? Qualitative or quantitative. If the answers are categories, no average will ever mean anything.

3. Counted or measured? Only for quantitative data. Counted is discrete, measured is continuous, and the reason matters as much as the word.

4. Tally, table, total. One mark per value, bundled in fives; one row per category; and the frequencies must add back to the number you started with.

THE MISTAKES THAT COST THE MOST MARKS

Trap 1 · Calling shoe size continuous because it is “a measurement”

Wrong. “Shoe size is measured off your foot, so it is continuous.”

Right. Shoe size is **discrete**. Sizes come in fixed steps — 5, 6, 7 and perhaps half sizes — and there is no size 6.31.

The foot inside the shoe *is* continuous; the size printed on the shoe is not. Ask what is actually written down, not what it came from. The same trap catches dress sizes and shoe widths.

Trap 2 · Forgetting the bundle of five, and miscounting

Wrong. Eighteen separate upright marks, counted one by one, and read off as 17.

Right. Bundle every fifth mark **across** the four before it, then count the bundles in fives: three bundles and three singles is 5 + 5 + 5 + 3 = 18.

The bundles are not decoration. Counting a long row of identical uprights is exactly where miscounts happen, and the whole table is wrong afterwards. The frequencies must add back to the number of values you started with — that is the check that catches it.

Trap 3 · Reading a two-way table cell as a row total

Wrong. “4 girls are in Form 1C, because the table says 4 under Girls.”

Right. The cell says **4 girls came by combi**. The number of girls altogether is the **row total**, which is 9.

A cell answers two questions at once and a total answers one. Say the cell out loud as a full sentence — “girls, and combi” — and the mistake becomes impossible to make.

PRACTICE · 60 QUESTIONS

Answers are at the back. Work each one out before you look.

Concept 1 · What counts as data

P1 (Fluency)	**State** the word for data you collected yourself.
P2 (Fluency)	**State** the word for data you took out of somebody else’s report.
P3 (Fluency)	Your class counts the vehicles at the school gate for twenty minutes. **Classify** the data as primary or secondary.
P4 (Fluency)	You copy last year’s rainfall figures out of a textbook. **Classify** the data as primary or secondary.
P5 (Applied)	You measure every learner’s hand span in your class. **Classify** the data and **describe** why.
P6 (Applied)	You use the results of a survey another class ran last term. **Classify** the data and **describe** one risk.
P7 (Applied)	“Way to school” is collected from 16 learners. **Describe** the difference between the variable and a data value here.
P8 (Applied)	You time your own walk to school every day for a week. **Classify** the data and **state** how many data values you have.
P9 (Challenge)	A learner says “anything from the internet is secondary data”. **Describe** a case where data from a website would be primary.
P10 (Challenge)	**Describe** one advantage of primary data and one advantage of secondary data.

Concept 2 · Words or numbers?

P11 (Fluency)	**Classify** the variable “favourite colour” as qualitative or quantitative.
P12 (Fluency)	**Classify** the variable “number of brothers and sisters” as qualitative or quantitative.
P13 (Fluency)	**Classify** the variable “home language” as qualitative or quantitative.
P14 (Fluency)	**Classify** the variable “shoe size” as qualitative or quantitative.
P15 (Applied)	**Classify** “eye colour” and **describe** the reason.
P16 (Applied)	**Classify** “temperature at noon” and **describe** the reason.
P17 (Applied)	A netball team records each player’s jersey number. **Describe** why this is qualitative data even though it is written with digits.
P18 (Applied)	**Classify** “day of the week” and **describe** the reason.
P19 (Challenge)	Four variables were collected: home language, number of pencils in the bag, favourite colour, height. **Classify** each one as qualitative or quantitative.
P20 (Challenge)	A learner writes the rule “if it is written with digits it is quantitative”. **Describe** why the rule fails, with an example.

Concept 3 · Counted or measured?

P21 (Fluency)	**Classify** “number of goals scored” as discrete or continuous.
P22 (Fluency)	**Classify** “height” as discrete or continuous.
P23 (Fluency)	**Classify** “number of pages in a book” as discrete or continuous.

P24 (Fluency)	**Classify** “mass of a loaf of bread” as discrete or continuous.
P25 (Applied)	**Classify** “shoe size” as discrete or continuous and **describe** the reason.
P26 (Applied)	**Classify** “time taken to get to school” and **describe** the reason.
P27 (Applied)	**Classify** “number of pencils in the bag” and **describe** the reason.
P28 (Applied)	**Classify** “length of a pencil” and **describe** the reason.
P29 (Challenge)	Three of these four were called discrete. **Identify** the one that is really continuous and **describe** why: number of learners in a class, mass of a loaf of bread, number of siblings, number of pages in a book.
P30 (Challenge)	**Describe** why the number of learners in a class can never be 27.5, and **state** what that tells you about the variable.

Concept 4 · Tally it, table it

P31 (Fluency)	Twenty vehicles are tallied. **State** how many tally marks the chart has altogether.
P32 (Fluency)	The traffic table is above. **State** the frequency of Combi.
P33 (Fluency)	The traffic table is above. **State** the frequency of Bakkie.
P34 (Fluency)	The traffic table is above. **State** the frequency of Lorry.
P35 (Applied)	**Calculate** the total of the five frequencies and **describe** what the total checks.

P36 (Applied)	**Calculate** how many of the twenty vehicles were *not* combis.
P37 (Applied)	**Calculate** how many more combis than bakkies went past.
P38 (Applied)	**Draw** the tally marks for a frequency of 13, and **describe** how you would count them back.
P39 (Challenge)	A learner's tally of the same twenty vehicles gives frequencies adding to 19. **Describe** what has gone wrong and what to do.
P40 (Challenge)	**Calculate** how many of the twenty vehicles had four wheels or more, using the frequency table.

Concept 5 · Two ways at once

P41 (Fluency)	The two-way table is above. **State** how many girls walk.
P42 (Fluency)	The two-way table is above. **State** how many boys come by combi.
P43 (Fluency)	The two-way table is above. **State** how many girls there are altogether.
P44 (Fluency)	The two-way table is above. **State** how many learners come by car.
P45 (Applied)	**Calculate** how many boys there are altogether, and **state** which numbers you added.
P46 (Applied)	**Calculate** how many learners are in Form 1C, two different ways.
P47 (Applied)	**Calculate** how many girls do *not* walk to school.
P48 (Applied)	**Calculate** how many more learners come by combi than by car.

P49 (Challenge)	A learner says “4 learners come by combi, because the table says 4 where Girls meets Combi”. **Describe** the error and **state** the correct number.
P50 (Challenge)	**Describe** the check that a two-way table is right, and **show** it for this table.

Concept 6 · Read it back

P51 (Fluency)	The sibling table is above. **State** how many learners have exactly 3 brothers and sisters.
P52 (Fluency)	The sibling table is above. **State** how many learners answered altogether.
P53 (Fluency)	The sibling table is above. **State** the most common number of brothers and sisters.
P54 (Fluency)	The sibling table is above. **State** how many learners have no brothers or sisters at all.
P55 (Applied)	**Calculate** how many learners have more than 3 brothers and sisters.
P56 (Applied)	**Calculate** how many learners do *not* have exactly 1 brother or sister.
P57 (Applied)	**Calculate** how many more learners have 1 brother or sister than have 4.
P58 (Applied)	The bar chart of the traffic count is above. **State** which bar is the tallest and **describe** what its height means.
P59 (Challenge)	**Describe** why the frequency axis of a bar chart must start at 0.
P60 (Challenge)	**Calculate** how many learners have fewer than 2 brothers and sisters, and **state** which rows you added.

CRITERION B · INVESTIGATING PATTERNS

Twenty questions

A sorting game. One player thinks of an object from a set laid out on the desk. The other player may only ask questions with a **yes** or **no** answer — “is it made of metal?”, “is it blue?” The question is how many objects a fixed number of questions can always tell apart.

Part 1 · the structured investigation

1. **Predict** how many different objects 5 yes/no questions could always tell apart. Write the prediction down before you work anything out.

2. Play the game with 1 question, then 2, then 3, and **tabulate** the largest number of objects that can always be told apart.

3. **Identify** the pattern in your table. Say what happens to the number each time one more question is added.

4. **Determine** the general rule for *n* questions, and write it as a formula.

5. **Test and verify** the rule: build the full question tree for 4 questions, count the branches at the end, and compare the count with your formula.

6. **Justify** why one more question can at most double the number of groups — not treble it, and not add two.

Number of questions, n	Objects that can always be told apart
1	2
2	4
3	8
4	
5	
n	

The first three rows are done for you as far as n = 3. The last three rows are yours.

Part 2 · your own method choice

Part 2 · your own investigation. Choose one of these, or an equally good question of your own, and investigate it in the same way — predict, tabulate, find the pattern, state a general rule, test it and justify it. (a) What happens when each question has *three* possible answers instead of two? (b) What is the smallest number of yes/no questions needed for a set of 30 objects, and how do you know it is the smallest? You choose the method, the table and the way you present the result.

Criterion B · Investigating patterns. Levels 1–2: applies the given procedure and records some correct values. **Levels 3–4:** completes the table correctly and describes the pattern in words. **Levels 5–6:** states a correct general rule for *n* questions and verifies it with the tree for 4. **Levels 7–8:** justifies the doubling from the structure of the tree — each existing group is split into a yes group and a no group and no group can split further than that — and carries Part 2 through to a tested general statement of its own.

Remember, you should communicate in an organized and coherent manner.

CRITERION D · APPLYING MATHEMATICS IN REAL-LIFE CONTEXTS

The tuck-shop record sheet

A tuck shop at a school in Gaborone keeps one line of a record sheet for every sale. The manager wants to know **what sells most at break time**, so that she orders the right amounts. Break runs from 10:15 to 10:45. The sheet carries more columns than the question needs.

Item	Price (P)	Time	Weather	Seller
Fat cake	5.00	07:50	sunny	MK
Juice	6.50	10:18	sunny	TB
Fat cake	5.00	10:20	sunny	TB
Chips	8.00	10:22	sunny	MK
Fat cake	5.00	10:24	sunny	MK
Apple	3.00	10:27	sunny	TB
Juice	6.50	10:30	sunny	MK
Fat cake	5.00	10:33	sunny	TB
Chips	8.00	10:36	sunny	MK
Sandwich	12.00	10:39	sunny	TB
Fat cake	5.00	10:42	sunny	MK
Juice	6.50	13:10	cloudy	TB
Chips	8.00	13:15	cloudy	MK
Fat cake	5.00	15:05	cloudy	TB

This record sheet is an invented illustrative example written for the task. It is not taken from any business, price list or publication, and no organisation, person or source is named or implied.

1. **Interpreting the situation. Describe** which columns of the sheet are needed to answer “what sells most at break time”, and which are surplus to that question. Say what each surplus column would be useful for instead.

2. **Selecting the mathematics. Classify** each of the five columns as qualitative or quantitative, and where it is quantitative, as discrete or continuous. Give the reason in each case.

3. **Applying the mathematics. Build** the tally chart and then the frequency table of the items sold *at break time only*. **State** which item sold most.

4. **Accuracy. Check** that your frequencies add to the number of rows you used, and **state** that number. Say how you decided which rows belonged to break and which did not.

5. **Sense in context. Describe** what the manager should order more of, and **describe** one thing this sheet cannot tell her — something she would need to collect differently to find out.

Criterion D · Applying mathematics in real-life contexts. Levels 1–2: identifies some of the needed columns and makes a start on the tally. **Levels 3–4:** selects the right columns, classifies most of them correctly and completes the frequency table for break time. **Levels 5–6:** all five classifications correct with reasons, the frequency table complete and checked against the row count, and a described recommendation. **Levels 7–8:** in addition, describes clearly what the sheet cannot answer — for example how many learners wanted something the shop had run out of, which no record of sales can ever show — and says what would have to be collected instead.

ANSWERS

Every answer here is recomputed from the datasets printed in the questions before this file is written.

P1	primary	**P31**	20
P2	secondary	**P32**	7
P3	primary	**P33**	4
P4	secondary	**P34**	1
P11	qualitative	**P41**	3
P12	quantitative	**P42**	3
P13	qualitative	**P43**	9
P14	quantitative	**P44**	4
P21	discrete	**P51**	5
P22	continuous	**P52**	24
P23	discrete	**P53**	1
P24	continuous	**P54**	3

Applied and Challenge · fully worked

P5 primary

You did the measuring, so the data is yours: it is primary. You can say exactly how it was collected, with what ruler, and on what day, which is the whole advantage of primary data.

P6 secondary

Somebody else collected it, so it is secondary. The risk is that you cannot check how they asked: who they asked, how many, and whether the question was fair. You are trusting their method as well as their numbers.

P7 the variable is “way to school”; a data value is one learner’s answer, such as “combi”

The variable is the thing being collected — it has one name and one meaning for the whole class. A data value is one single answer from one single learner. Sixteen learners give one variable and sixteen data values.

P8 primary, and 7 data values

You collected it yourself, so it is primary. One timing per day for seven days is seven data values of the one variable “time taken to walk to school”.

P9 if your own class collected it and put it on the website, it is still primary

Primary and secondary are about *who collected the numbers*, not about where they are stored. If your class ran the traffic count and then typed the results onto a class web page, the data is primary for you and secondary for anyone else who downloads it. The same numbers can be primary for one person and secondary for another.

P10 primary: you know exactly how it was collected; secondary: it is much quicker and can cover far more people

Primary data is trustworthy because you controlled the method — you know who was asked and how. Secondary data is cheap and fast, and can reach a scale you could never collect yourself, such as a whole country. Neither is better in general; they answer different sized questions.

P15 qualitative

The answers are words — brown, black, hazel. There is no arithmetic you could do with them: brown plus black is not a colour, and there is no order from smallest to largest. That is exactly what qualitative means.

P16 quantitative

The answers are numbers on a scale — 31°C, 28°C — and arithmetic on them means something: one day really can be 3 degrees warmer than another.

P17 the digits are labels, not amounts

Jersey 7 is not “one more” than jersey 6, and the average jersey number of a team means nothing at all. The digits name a player the way a word would, so the variable behaves as a category. The test is never whether digits appear; it is whether arithmetic on them means anything.

P18 qualitative

The answers are the names of days. They do have an order, but they are still categories: Tuesday plus Thursday is not a day, and there is no sensible average day.

P19 qualitative: home language, favourite colour; quantitative: number of pencils in the bag, height

Two of them give words — a language and a colour — and two give numbers you can add and compare. Sorting first, before any chart is chosen, is what stops a pie chart of heights or a mean favourite colour.

P20 it fails for qualitative labels written as digits, such as a jersey number or a house number, which are never quantitative

House number 42 is not twice house number 21, and the two cannot be added to give anything. Digits are being used as a name. The working rule is: if arithmetic on the values means something, the variable is quantitative; if the values only name a category, it is qualitative.

P25 discrete

Sizes come in fixed steps — 5, 6, 7, and half sizes where the shop stocks them. There is no size 6.31, so the values are separate and countable. The foot inside the shoe is continuous; the size printed on it is not.

P26 continuous

Time is measured, not counted. Between 12 minutes and 13 minutes every value is possible, and a stopwatch with more decimal places just finds more of them. Rounding the answer to the nearest minute does not make the variable discrete — it only makes the record rougher.

P27 discrete

Pencils are counted one at a time, so the answers are whole numbers with nothing in between. A broken pencil is still one pencil, or none.

P28 continuous

A pencil can be any length at all. Measuring it more carefully adds decimal places rather than landing on a new fixed value, and that is exactly what continuous means.

P29 mass of a loaf of bread

Learners, siblings and pages are all counted in whole numbers, so all three are discrete. Mass is read off a scale: a loaf can be 748 g or 748.3 g or anything between, so it is continuous. The test that finds the impostor every time is “would I reach for a measuring instrument?”

P30 learners are counted in whole numbers, so the variable is discrete

A class has 27 learners or 28 learners; there is no learner-and-a-half to be had. Values that can only land on separated whole numbers are exactly what discrete means. Note that an *average* class size of 27.5 is perfectly sensible — the average is a calculation, not a data value.

P35 20

7 + 5 + 4 + 1 + 3 = 20, which is the number of vehicles in the list. The total is the check: if the frequencies had come to 19 or 21, a mark has been lost or double-counted and the table cannot be used.

P36 13

Total minus that one frequency: 20 − 7 = 13. Adding the other four frequencies gives the same answer, which is a useful check.

P37 3

“How many more” is a subtraction of the two frequencies, never an addition: 7 − 4 = 3. Say the answer as a sentence — there were 3 more combis than bakkies — because that is what stops the subtraction being done the wrong way round.

P38 |||| |||| |||

Two full bundles of five and three singles. You count them back in fives: 5, 10, then the three singles, giving 13. Bundling is what makes the count in fives possible, and counting in fives is what stops the miscount.

P39 one vehicle has been missed; go back to the raw list and tally again

The frequencies must add back to the number of values in the list. Nineteen against a list of twenty means exactly one mark was never made, or one row was miscounted. There is no way to repair it from the table alone — the raw list is the only evidence, so the tally is done again against it.

P40 17

Only the motorbikes have fewer than four wheels, so this is the total minus that frequency: 20 − 3 = 17. The same answer comes from adding combi, car, bakkie and lorry: 7 + 5 + 4 + 1 = 17.

P45 7

Add the whole Boys row: 2 + 3 + 2 = 7. A row total answers a question about the side variable only — here, gender.

P46 16

By rows: 9 + 7 = 16. By columns: 5 + 7 + 4 = 16. The two routes must agree, because every learner sits in exactly one cell and is therefore counted exactly once each way.

P47 6

The girls' row total minus the walking cell: 9 − 3 = 6. Those are the girls who came by combi or by car.

P48 3

Two column totals, subtracted: 7 − 4 = 3. Note that both numbers are totals, not cells — the question is about the whole form, not about girls or boys separately.

P49 7

The cell 4 answers two questions at once: girls, *and* combi. "How many come by combi" asks about one variable only, so it is the column total: 4 + 3 = 7.

P50 the row totals and the column totals must give the same grand total: 16

Rows: 9 + 7 = 16. Columns: 5 + 7 + 4 = 16. They agree, so every learner has been counted exactly once. If they disagreed, at least one learner has been placed in two cells or in none, and the table has to be rebuilt from the raw list.

P55 3

"More than 3" means 4 or 5, so add those two rows: 2 + 1 = 3. It does not include the row for 3 itself.

P56 17

Total minus that one frequency: 24 − 7 = 17. Adding the other five rows gives the same number, which is the check.

P57 5

A "how many more" question subtracts the two frequencies: 7 − 2 = 5. Both numbers are frequencies — how many learners — and not numbers of brothers and sisters, and saying the answer as a sentence keeps the two apart.

P58 Combi, at a height of 7

The tallest bar is the one with the largest frequency. Its height is read off the frequency axis as 7, and that means 7 combis went past the gate in the twenty minutes — not 7 minutes, and not 7 anything else. A bar's height is always a frequency.

P59 so that each bar's height stays in proportion to its frequency, which only happens when the axis starts at 0

If the axis started at, say, 4, the bar for a frequency of 5 would be one unit tall and the bar for 7 would be three units tall — the picture would say "three times as many" when the truth is "a bit more". Starting at 0 is what makes the comparison you see the comparison that is true. Unit S03 is about charts that break this rule on purpose.

P60 10

"Fewer than 2" means 0 or 1, so the rows for 0 and 1 are added: 3 + 7 = 10. The row for 2 is not included — "fewer than" excludes the value itself, which is the commonest slip in this kind of question.