nexamind EDUCATION

Name _____ Date _____

Unit 6 · Collecting data — teaching slides

12 slides, formatted for printing.

6.1 Conducting an investigation

What makes a question statistical

- A statistical question has an answer that **varies**.
- "How tall am I?" is not statistical — one measurement answers it.
- "How do heights vary in my class?" is — you need data.
- If you can answer it without collecting anything, it is not one.

```
✗ What is my shoe size?
✓ What is the most common shoe size in Year 7?
✓ Do learners who walk to school arrive earlier?
```

Plan before you collect

- Write the question.
- Write a **prediction** — what you think you will find.
- Decide what data will actually test that prediction.
- Only then start collecting. Collecting first wastes the effort.

```
Question: which sport is most popular in Year 7?
Prediction: football, because most clubs are football clubs
Data needed: each learner's favourite sport (categorical)
```

Three types of data

- **Categorical** — words or labels: eye colour, country.
- **Discrete** — numbers you **count**: goals, pets, siblings.
- **Continuous** — numbers you **measure**: height, time, mass.
- The test: could the value be 3.7? If yes it is continuous.

```
favourite colour → categorical
number of pets → discrete
height in cm → continuous
```

Tally charts and frequency tables

- Tally as the data arrives — the five-bar gate is easy to count.
- The tally totals become the **frequency**.
- Add a Total row and check it against how many people you asked.
- If the total is wrong, the count is wrong. Check before you draw anything.

```
dog 𝍸 || 12
cat 𝍸 |||| 9
fish 𝍸 5
Total 26
```

Grouping continuous data

- Measurements are nearly all different, so a plain tally is useless.
- Sort them into **class intervals** instead.
- The groups must not overlap, and must leave no gaps.
- Write them with inequalities so every value has exactly one home.

```
✗ 140–150, 150–160 where does 150 go?
✓ 140 ≤ h < 150
✓ 150 ≤ h < 160
```

Asking a fair question

- A **leading** question pushes people towards one answer.
- Answer boxes must cover every case, exactly once.
- Keep it short, and ask one thing at a time.
- Test it on two people before you use it on two hundred.

```
✗ Don't you agree football is the best sport?
✓ Which sport do you most enjoy watching?
✗ 0–5, 5–10, 10–15 (5 and 10 fit twice)
✓ 0–4, 5–9, 10–14
```

6.2 Taking a sample

Population and sample

- The **population** is every member of the group you are studying.
- It need not be a country — it might be Year 7 at your school.
- A **sample** is the part you actually collect data from.
- Asking everyone is a **census**: most reliable, rarely possible.

```
Population: all 800 learners in the school
Sample: 40 of them
Census: all 800 — accurate, but a whole day's work
```

Why sample size matters

- A small sample is easily swayed by one unusual answer.
- With 10 people, one extra yes moves the result 10 percentage points.
- With 500, one answer moves it by 0.2.
- Bigger samples give more reliable results — but never exact ones.

```
10 asked → 7 yes → 70%
one more yes → 80% a huge jump
500 asked → 260 yes → 52%
one more yes → 52.2% barely moves
```

Three ways to choose

- **Random** — every member has an equal chance. The gold standard.
- **Systematic** — a fixed rule: every 10th name on the register.
- **Convenience** — whoever is easiest to reach. Fast, and usually biased.
- Ask yourself: could anyone in the population have been chosen?

```
names from a hat → random
every 5th on the list → systematic
my friends at break → convenience
```

Bias

- A sample is **biased** when some members are more likely to be chosen.
- The result is then misleading no matter how large the sample is.
- Look at *where* and *when* you sampled, and *who* could reply.
- This is the single most important idea in the section.

```
at a bus stop → only bus users
the football club → only football fans
online only → only people with internet
phone-in → only people who feel strongly
```

Size does not fix bias

- These are two separate problems, and they need two separate fixes.
- Too small → collect more data.
- Biased → change **how** you choose, not how many.
- A biased sample of 10 000 is still biased.

```
20 football fans asked → football wins
200 football fans asked → football still wins
20 random learners asked → a fair picture
```

From sample to population

- Find the proportion in the sample.
- Apply that proportion to the whole population.
- Say clearly that it is an **estimate**, and give the sample size.
- The bigger and fairer the sample, the better the estimate.

```
24 of 60 play an instrument
24 ÷ 60 = 0.4
0.4 × 900 = 360 learners, estimated
```