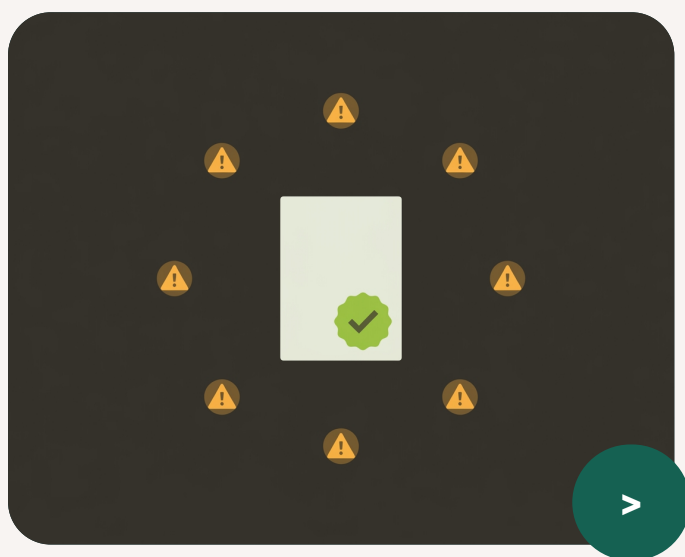


Sampling and field data, without the maths degree.

How many people to survey, how to choose them, and how to trust the data once it arrives. Plain English, with the assumptions stated — so you can defend your sample and clean your dataset before anyone questions it.



What you will be able to do

Choose a sample size and a design, state the assumptions, run five data-quality checks, and clean a dataset so a donor can trust every number.

■ PART ONE

How many people do you need?

For a survey estimating a proportion, the answer depends on three things and nothing else. State them and the number is defensible.

95%

Confidence level — the default. How sure you want to be.

$\pm 5\%$

Margin of error — how close the estimate must be.

The third assumption

The expected proportion. Use 0.5 unless you already know better — it is the safe, conservative default.

■ IN SUMMARY

Sample size is not a guess. It is three assumptions and a formula.

The numbers, in a table

For a proportion at 95% confidence and $\pm 5\%$ margin, the size depends on the population only when it is small.

10,000+	~384 for $\pm 5\%$ at 95%
1,000	~278
300	~169
100	~80

WHAT IS "POPULATION"?

Everyone you could survey — all the households, schools or beneficiaries in scope.

WHAT IS "SAMPLE"?

The subset you actually survey. You use it to say something about the population.

Choosing who: the designs

How you choose matters as much as how many. State the design in your methods section.

Simple random

Everyone has an equal chance. Strongest, but often impractical in the field.

Cluster

Sample villages or schools, then people within them. Practical; weaker — multiply the sample by the design effect (often 1.5–2).

Purposive

Chosen deliberately for depth — for qualitative work, not for prevalence.

If you cannot state the design, you cannot defend the finding.

Five checks before you analyse anything

Most survey analysis is wrong before it starts. Run these five checks and log every exclusion.

Check 1	Completeness
Check 2	Duplicates
Check 3	Outliers
Check 4	Skip-logic violations
Check 5	Enumerator patterns

The single most common error is treating a blank as a zero — which quietly deflates every average. A blank is not a zero.

Mobile data collection that works

Kobo and ODK are the two tools most teams use. The failures are almost always the same, and almost always avoidable.

1

Pre-load the forms

Before fieldwork, not on the first morning.

2

Confirm device date and time

Wrong clocks corrupt every timestamp.

3

Test skip logic

One dummy submission catches most form errors.

4

Sync over Wi-Fi

Mobile data drops mid-upload and creates duplicates.

Cleaning the data, reversibly

Cleaning is a fixed sequence. Keep it reversible and documented.

- 1 Freeze a raw copy**
Never edit the original export.
- 2 Standardise**
Text and dates into one format.
- 3 Resolve duplicates**
Decide and record which record survives.
- 4 Recode missing vs zero**
Keep them distinct.
- 5 Check ranges and logic**
Flag anything impossible.
- 6 Document every change**
A log, so any decision can be reversed.

Baselines when there is none

No baseline is not a dead end. The honest options, best first.

- Reconstruct from existing administrative records.
- Run a rapid baseline now, dated honestly.
- Measure a comparison group today.
- Report from the first measurement onward — and say so.

“No baseline; first measurement October 2026.” Honesty is cheaper than a sanction.

Five mistakes to avoid

Each is common, and each is fixable.

- A sample size with no stated assumptions.
- A blank treated as a zero.
- Cluster design with no design effect.
- Cleaning that cannot be reversed.
- An unexplained sample, which invites doubt about every number built on it.

■ GO FURTHER — THE PRO EDITION

The Pro edition adds the sample-size calculator sheet, the data-quality workbook, the cleaning log, and the Kobo/ODK checklists.

\$79 · downloadable PDF with every template. Or get it included with the video course. For teams, write to contact@evaluationai.com.