THE GRIPHAN ONLINE SCHOOL

UNIVERSITY LIBRARY

Statistics for Researchers I

Describing data honestly: first steps for every researcher

Nji Griphan

griphan.48hourrfprescue.com

Statistics for Researchers I

Describing data honestly: first steps for every researcher

Nji Griphan

The Griphan Online School — University Library

IMPRINT

This book is published for the Library of the Griphan Online School. Author and publisher of record: **Nji Griphan**. School address on the web: **griphan.48hourrfprescue.com**. Professional mailbox: **info@griphan.48hourrfprescue.com**. © 2026 Nji Griphan. This book was written for the students of this school. Students enrolled at the school may download and keep it for their own study. The library grows shelf by shelf until it carries the whole university.

A note on this book

Statistics is the grammar of research evidence. This first volume teaches you to describe data truthfully — averages, spread, tables, charts and proportions — and to avoid the classic deceptions, including the ones researchers practise on themselves. No prior mathematics beyond school arithmetic is assumed; every idea is built on worked examples from field and social research.

Contents

Chapter 1 — Data — What We Actually Record

Chapter 2 — Tables and Charts That Tell the Truth

Chapter 3 — Averages — Mean, Median and Mode

Chapter 4 — Spread — How Far Values Scatter

Chapter 5 — Proportions, Ratios and Rates

Chapter 6 — Reading Studies Critically

Appendix — Worked Examples and Practice

Glossary of Key Terms

Chapter 1

Data — What We Actually Record

Before any calculation, a researcher must know what kind of thing each recorded value is. This chapter classifies data and shows how the kind of data decides every later choice of chart, average and test.

Variables and observations

An observation is one measured unit — a plot, a household, a student. A variable is what you record about it: yield, income, examination score. Keep one row per observation and one column per variable in a tidy table; nearly every later confusion in analysis begins as a messy table at collection time.

Kinds of variables

Categorical variables name groups — village, sex, crop variety. Ordered categories carry rank without equal steps — small, medium, large farms. Numerical variables carry amounts — kilograms, years, naira — and may be counts (whole numbers) or measurements (any value in a range). Name each variable's kind in your notebook before analysis day.

Measurement and its errors

Every measurement carries error: the scale drifts, the respondent rounds, the enumerator mishears. Distinguish random error, which averages out over many observations, from systematic error — bias — which pushes every value the same wrong way and no sample

size can cure. Design against bias first.

Recording and storing data

Record in the field book first, enter into the computer second, and keep the two reconcilable by matching identification numbers. Back up in two places. A dataset whose origin cannot be traced is an opinion with decimals.

Key points

- One row per observation, one column per variable.
- The kind of variable decides the chart, average and test.
- Sample size cures random error; only design cures bias.

Exercises

1. Classify: crop variety; yield in kg; farm size class (small/medium/large); number of children.

2. Give one example each of random and systematic error in a yield survey.

3. Design the column headings for a tidy table of a variety trial.

Chapter 2

Tables and Charts That Tell the Truth

A table or chart is an argument in visual form, and it can argue falsely. This chapter teaches the standard displays — frequency tables, bar charts, histograms, pie charts and line charts — and the honesty rules that govern them.

Frequency tables

A frequency table counts how many observations fall in each category or interval, and adds relative frequencies (proportions or percentages) so groups of different sizes can be compared. Always print the total count beside the percentages; a percentage without its base is a claim without evidence.

Bar charts and histograms

Bar charts compare categories with separated bars; histograms show the shape of numerical data with joined bars over equal intervals. Both live or die by their axes: a bar chart whose axis does not start at zero exaggerates differences and deceives the eye, however accurate its numbers.

Pie charts and their limits

A pie chart shows parts of one whole. It fails when slices are many or similar in size, because eyes judge angles badly. If a bar chart would show it more clearly — it usually would — use the bar chart. Reserve pies for two to four parts of a single whole.

Line charts for change over time

Plot time on the horizontal axis and the measured value vertically, joining points in order. Line charts reveal trends, seasons and shocks — the drought year stands out at a glance. Never join points across categories; a line implies a path through time that categories do not have.

Key points

- Percentages must travel with their counts.
- Bar axes start at zero; always.
- Choose the chart by the question, not by decoration.

Exercises

1. Build a frequency table from 30 invented household sizes, with percentages.

2. A chart's axis starts at 90 to show values 95 and 100. Explain the deception and redraw honestly in words.

3. Choose chart types for: rainfall by month; households by main crop; yields of 40 plots.

Chapter 3

Averages — Mean, Median and Mode

An average is a promise that one number can stand for many. This chapter teaches the three common averages, when each keeps its promise, and when each breaks it.

The mean

The mean adds all values and divides by their count. It uses every observation, which is its strength, and it is dragged by extreme values, which is its weakness. One wealthy household can pull a village's mean income far above what most households earn.

The median

The median is the middle value when observations are ranked. Half lie above it, half below. It ignores extremes, so it describes the typical household, worker or plot better whenever the distribution is skewed — as incomes, farm sizes and land holdings almost always are.

The mode

The mode is the most frequent value — the commonest household size, the variety most farmers plant. For categorical data it is the only average available. A distribution may have two modes, which itself is information: two distinct groups may be hiding inside your data.

Choosing the average

Ask what question the average must answer. For totals and further calculation, the mean; for the typical case under skew, the median; for the commonest category, the mode. Honest reports of skewed data give mean and median together and let their gap speak.

Key points

- Mean: uses all values, dragged by extremes.
- Median: the typical value, robust under skew.
- Report mean and median together for skewed data.

Exercises

1. Compute mean and median of: 2, 3, 3, 4, 5, 6, 40 (hectares). Which describes the typical farm?

2. Give a research question for which the mode is the right summary.

3. A mean income exceeds the median by half. What does the gap tell you?

Chapter 4

Spread — How Far Values Scatter

Two datasets can share an average and differ in everything that matters. This chapter measures scatter — range, interquartile range and standard deviation — and teaches you to report averages and spread together, always.

The range

The range, largest minus smallest, is the quickest spread measure and the most fragile, since two extreme values decide it entirely. Use it as a first glance and a data-checking tool: a range beyond the possible — a household of 90 persons — flags a recording error.

Quartiles and the interquartile range

Quartiles divide ranked data into quarters. The interquartile range (IQR), from the first to the third quartile, holds the middle half of the data and, like the median it accompanies, resists extremes. Report median with IQR as you report mean with standard deviation.

The standard deviation

The standard deviation measures the typical distance of values from their mean, in the original units. Small beside the mean: values cluster. Large: they scatter. It is the spread measure of scientific reporting, and the raw material of the tests in Volume II.

Why spread is a finding

Spread is often the story: a variety with a slightly lower mean but half the spread may serve a poor household better than a brilliant but erratic one, because the family eats the bad year too. Never let a report of averages travel without its spread.

Key points

- Averages without spread are half a report.
- Median pairs with IQR; mean pairs with standard deviation.
- Low spread can be worth more than a high mean.

Exercises

1. Compute the range and IQR of: 4, 6, 7, 8, 9, 11, 15.

2. Two varieties both average 20 bags; spreads differ threefold. Which serves a risk-averse household?

3. Explain why a range of 0–90 for household size should trigger a data check.

Chapter 5

Proportions, Ratios and Rates

Much research speaks in proportions — the share of households food insecure, the germination rate, the proportion of variance explained. This chapter builds these measures carefully, because small confusions here produce large public errors.

Proportions and percentages

A proportion divides a part by its whole: 46 of 200 households = 0.23, or 23 per cent. The whole must be stated and must be the right whole — the proportion of women among farmers is not the proportion of farmers among women, and confusing the two has misled real policies.

Ratios

A ratio compares two quantities that need not be part and whole: 30 pupils per teacher, 2.4 hectares per household. Ratios compare across places and times fairly only when their definitions match exactly; check definitions before comparing any two published ratios.

Rates and time

A rate ties events to exposure over time — births per 1,000 people per year, infections per 100 plant-weeks. The time dimension is not decoration: a count of cases without its period and population cannot be compared with anything, including itself next year.

Percentage points versus per cent

A rise from 20 per cent to 25 per cent is a rise of 5 percentage points, and also a 25 per cent relative increase. Both are true; they create opposite impressions. Honest writing states which is meant, every time.

Key points

- State the whole behind every proportion.
- Rates need population and period.
- Distinguish percentage points from per cent change.

Exercises

1. Of 240 seedlings, 204 survive. Give the survival proportion and percentage.

2. Unemployment moves from 10% to 12%. State the change both ways.

3. Why is '30 cases' incomparable between a village and a city? Repair the comparison.

Chapter 6

Reading Studies Critically

The final chapter of Volume I turns the tools backward onto published work — your own included. You will learn the standard questions that expose weak description before it hardens into accepted fact.

Who was measured?

Find the population, the sample and the response rate. A survey of 'farmers' that interviewed only those at a cooperative meeting describes cooperative attenders. If the paper cannot tell you who was measured and who refused, its conclusions float free of any population.

What exactly was counted?

Definitions decide results: 'food insecure' by which scale, 'yield' weighed or estimated, 'employed' by which hours threshold? Compare definitions before comparing numbers between studies; two honest studies can disagree purely because their rulers differ.

Where are the spreads and counts?

Apply this volume: percentages without bases, means without spreads, charts with wounded axes. Each omission you can now name is a place where the argument may be hiding weakness — or simple haste. Judge accordingly, and correct it in your own writing first.

Correlation is not causation

Two things rising together — ice-cream sales and drowning — may share a cause (summer) without causing each other. Descriptive statistics can establish that variables move together; claiming one moves the other needs design, the subject of Volume II. Guard this line in every conclusion you write.

Key points

- Ask: who was measured, what was counted, where are counts and spreads?
- Definitions decide results; compare them before numbers.
- Description shows association, not causation.

Exercises

1. Take any news report of a study and list what it omits from this chapter's questions.

2. Give an example of two correlated variables with a shared third cause.

3. Rewrite a causal headline as an honest descriptive one.

Appendix — Worked Examples and Practice

From Chapter 1 — Data — What We Actually Record

Worked example: 'yield' recorded as 'about two bags' mixes units and estimation into one cell. The tidy version records three columns — number of bags (2), weight per bag in kg (48), and whether weighed or estimated (estimated) — so the analyst can compute 96 kg and flag the value as an estimate. Detail at recording saves the analysis.

From Chapter 2 — Tables and Charts That Tell the Truth

Worked example: two groups score 95 and 100 on an index. On an axis starting at zero their bars look nearly equal — correctly, since the difference is 5 per cent. Truncate the axis at 90 and the second bar looks twice the first. The numbers never changed; the lie is in the axis. Examiners and editors reject truncated bar axes for this reason.

From Chapter 3 — Averages — Mean, Median and Mode

Worked example: farm sizes 2, 3, 3, 4, 5, 6, 40 hectares sum to 63, so the mean is 9 ha — larger than six of the seven farms. The median is 4 ha, the true middle farm. Reporting only the mean would describe a village of comfortable farms that does not exist; the median tells the truth about the typical holding.

From Chapter 4 — Spread — How Far Values Scatter

Worked example: values 4, 6, 7, 8, 9, 11, 15 have median 8. The lower half (4, 6, 7) has median 6 = Q1; the upper half (9, 11, 15) has median 11 = Q3. The IQR is 11 − 6 = 5: the middle half of the data lies within a band of 5 units, a statement no single outlier can falsify.

From Chapter 5 — Proportions, Ratios and Rates

Worked example: a pass rate rises from 40 per cent to 50 per cent. The absolute change is 10 percentage points. The relative change is (50−40)/40 = 25 per cent. A headline crying 'passes up 25 per cent' and one reporting 'up 10 points' describe the same classrooms; the researcher must give both numbers and let readers judge.

From Chapter 6 — Reading Studies Critically

Worked example: 'villages with more extension visits have higher yields, so visits raise yields.' The description may be accurate, yet keen farmers may both seek visits and farm better — the visits could be effect as much as cause. The honest sentence is the descriptive one; earning the causal one requires the designs of Volume II.

Further Practice — Review and Discussion

Work through these prompts in writing; they are the revision spine of this book. Answer in full sentences, with examples from your own place of study wherever the prompt allows.

Chapter 1 — Data — What We Actually Record

- Explain in your own words, with one example: One row per observation, one column per variable.
- Explain in your own words, with one example: The kind of variable decides the chart, average and test.
- Explain in your own words, with one example: Sample size cures random error; only design cures bias.

Discussion: where does the subject of this chapter — data — what we actually record — touch your own community, farm, school or workplace? Write one page, naming real places and real constraints, and end with one question you would research next.

Chapter 2 — Tables and Charts That Tell the Truth

- Explain in your own words, with one example: Percentages must travel with their counts.
- Explain in your own words, with one example: Bar axes start at zero; always.
- Explain in your own words, with one example: Choose the chart by the question, not by decoration.

Discussion: where does the subject of this chapter — tables and charts that tell the truth — touch your own community, farm, school or workplace? Write one page, naming real places and real constraints, and end with one question you would research next.

Chapter 3 — Averages — Mean, Median and Mode

- Explain in your own words, with one example: Mean: uses all values, dragged by extremes.
- Explain in your own words, with one example: Median: the typical value, robust under skew.
- Explain in your own words, with one example: Report mean and median together for skewed data.

Discussion: where does the subject of this chapter — averages — mean, median and mode — touch your own community, farm, school or workplace? Write one page, naming real places and real constraints, and end with one question you would research next.

Chapter 4 — Spread — How Far Values Scatter

- Explain in your own words, with one example: Averages without spread are half a report.
- Explain in your own words, with one example: Median pairs with IQR; mean pairs with standard deviation.
- Explain in your own words, with one example: Low spread can be worth more than a high mean.

Discussion: where does the subject of this chapter — spread — how far values scatter — touch your own

community, farm, school or workplace? Write one page, naming real places and real constraints, and end with one question you would research next.

Chapter 5 — Proportions, Ratios and Rates

- Explain in your own words, with one example: State the whole behind every proportion.
- Explain in your own words, with one example: Rates need population and period.
- Explain in your own words, with one example: Distinguish percentage points from per cent change.

Discussion: where does the subject of this chapter — proportions, ratios and rates — touch your own community, farm, school or workplace? Write one page, naming real places and real constraints, and end with one question you would research next.

Chapter 6 — Reading Studies Critically

- Explain in your own words, with one example: Ask: who was measured, what was counted, where are counts and spreads?
- Explain in your own words, with one example: Definitions decide results; compare them before numbers.
- Explain in your own words, with one example: Description shows association, not causation.

Discussion: where does the subject of this chapter — reading studies critically — touch your own community, farm, school or workplace? Write one page, naming real places and real constraints, and end with one question you

would research next.

Glossary of Key Terms

Variable — A characteristic recorded for each observation.

Bias — Systematic error pushing values consistently in one direction.

Median — The middle value of ranked data.

Standard deviation — The typical distance of values from their mean.

Interquartile range — The spread of the middle half of ranked data.

Proportion — A part divided by its whole.

Rate — Events relative to population and time.

Correlation — The tendency of two variables to move together, without proof of causation.

End of book. Nji Griphan · griphan.48hourrfprescue.com · info@griphan.48hourrfprescue.com