research nest

LESSON 6 · RESEARCH NEST COURSE

Collecting and Analysing Evidence

> *Noisier study spaces go with lower concentration ratings. Does that mean noise causes poor concentration, or could something else be going on?*

Sampling and data quality

Numbers and words: two kinds of analysis

Uncertainty, bias and correlation versus causation

LEARNING OBJECTIVES

What you will be able to do today

By the end you will have a one-page analysis plan for your own project.

- Judge how a sample and its data quality limit what you can claim
- Choose descriptive statistics and qualitative coding that fit your question
- Report association with uncertainty, and name likely confounders

1 **Raw data**
Survey rows, readings, transcripts, archival notes

→

2 **Clean**
Fix or flag errors, missing values, duplicates

→

3 **Describe**
n, mean or median, spread; first codes

→

4 **Analyse**
Compare groups, correlation, themes

→

5 **Interpret**
Cautious claims plus limitations

CONCEPT 1 · SAMPLING AND DATA QUALITY

Who and what is in your sample matters

A sample is the part of the population you measure. How it was chosen limits who your findings describe.

Good data are accurate, complete and measured the same way each time

Common sampling approaches

Approach	How it works	Main risk
Simple random	Everyone has an equal chance of selection	Hard to arrange in a school
Stratified	Random picks within groups (e.g. year levels)	Needs a list of each group
Convenience	Whoever is easy to reach	May not represent the population
Self-selected	People choose to take part	Volunteers may differ from others

Adapted from: OpenStax Intro Statistics 2e §1.2; ABS, Types of error

CONCEPT 2 · QUANTITATIVE AND QUALITATIVE

Counting and interpreting: two ways to analyse

Quantitative analysis summarises how much or how often; qualitative analysis explores meaning.

For 1–5 ratings, also give the median: scale steps may not be equal

Mixed methods: numbers show the pattern; interviews suggest why

QUANTITATIVE

Start with n. Describe the centre (mean or median) and spread (range or standard deviation). Then compare groups or test for association.

QUALITATIVE

Read and re-read. Label meaningful bits with short codes. Group codes into themes that answer your question. Keep quotes; note cases that do not fit.

Adapted from: OpenStax Intro Statistics 2e §2.7; Braun & Clarke, Understanding TA

CONCEPT 3 · INTERPRETING HONESTLY

Uncertainty, bias and the causation trap

Correlation does not, by itself, show causation

More data reduce random error, not bias

Uncertainty

Every measurement and sample has error. A water-test kit reading of 5 mg/L might really be 4.5–5.5 mg/L.

Bias

A systematic lean in one direction: a leading question, a self-selected sample, a meter that always reads high.

Correlation

A number from −1 to +1 showing the direction and strength of an association between two variables.

Confounding

A third variable linked to both. Hot weather lifts ice-cream and sunscreen sales; neither causes the other.

Adapted from: ABS, Correlation and causation; ONS, Uncertainty; OpenStax §12.3

WORKED EXAMPLE · STEP 1 INVENTED EXAMPLE · SIMULATED DATA

Step 1: Describe each study space

Mean noise rises from 46 to 65 dB(A) as mean concentration falls from 3.9 to 2.6

Group sizes are similar (26–31), so no space dominates the totals

Next: add medians and spread before drawing conclusions

Mean self-reported concentration (1–5) by study space, 84 responses over 2 weeks (simulated)



WORKED EXAMPLE · STEP 2 INVENTED EXAMPLE · SIMULATED DATA

Step 2: Read rho = −0.48 with care

NOOR'S DRAFT RESULTS SENTENCE, ANNOTATED

Across 84 study sessions, higher measured noise was associated with lower self-reported concentration [1] (Spearman's rho = −0.48), [2] a moderate negative association in this simulated dataset. [3] Because students chose their own spaces and noise readings were approximate, this does not show that noise caused lower concentration. [4]

States direction and n; says "associated", not "caused"

Negative sign: as one rises, the other tends to fall

Size: clear but far from perfect (−1)

Names confounding and measurement limits

WORKED EXAMPLE · STEP 3 INVENTED EXAMPLE · SIMULATED DATA

Step 3: From interview quotes to themes

Codes stay close to the words; themes answer the question

Theme 2 helps explain the self-selection confounder

Coding six short interviews (invented quotes)

Quote (invented)	Code	Theme
"Conversations pull me in; the fan I don't notice."	Speech vs steady noise	1. Unpredictable speech distracts most
"Someone laughing at the next table kills my focus."	Sudden voices	1. Unpredictable speech distracts most
"I sit in the cafeteria because my friends are there."	Friends decide location	2. Social reasons for louder spaces
"With headphones on, the cafeteria is fine for maths."	Headphones block noise	3. Headphones as a coping strategy

FRAMEWORK

Can I claim causation? A decision guide

Most student projects land on "association plus limitations": a valid result

Your two variables are associated. What may you claim?	
You randomly assigned conditions and controlled others	→ A causal claim may be reasonable for that setting
People chose their own groups (observational data)	→ Claim an association; name likely confounders such as self-selection
Measurements are approximate or self-reported	→ Add a measurement limitation and avoid precise-sounding claims
Small sample, one site or a short time window	→ Describe findings as specific to that setting, not universal

Adapted from: OpenStax §1.4, §12.3; ABS, Correlation and causation

COMMON MISTAKES

Common analysis mistakes and how to fix them

Mistake	Fix
"Noise causes poor concentration."	Say "was associated with" and name a likely confounder
Reporting only means	Add n, median and spread (range or standard deviation)
Silently deleting odd data	Write cleaning rules first; report what was removed and why
Cherry-picking quotes that fit	Code all transcripts; report cases that do not fit
Treating app readings as exact	Call them approximate; use one phone and one method

Adapted from: Purdue OWL, Analyzing your primary data; ABS, Types of error

GUIDED PRACTICE · 15 MINUTES · PAIRS INVENTED EXAMPLE · SIMULATED DATA

Practice: Clean, describe and claim

1. Find three problems and write a cleaning rule for each (5 min)
2. Rewrite "Cafeteria noise ruins concentration" as a cautious claim (5 min)
3. Name one confounder and one way to reduce it next time (5 min)

Practice extract of five raw survey rows (simulated). Spot the data-quality problems.

ID	Space	Noise dB(A)	Rating (1–5)
S01	Library	45	4
S02	Cafeteria	67	7
S03	Study hall	53	(blank)
S04	Cafeteria	64	2
S04	Cafeteria	64	2

MODEL RESPONSE INVENTED EXAMPLE · SIMULATED DATA

Model: Noor's results and limitations paragraph

DRAFT PARAGRAPH WITH ANNOTATIONS

Mean concentration ratings were highest in the library quiet zone (3.9/5, n = 31) and lowest in the cafeteria corner (2.6/5, n = 26).[1] Session noise and concentration showed a moderate negative association (Spearman's rho = −0.48).[2] Interviews suggested unpredictable speech distracted most, some chose louder spaces for friends, and headphones helped.[3] As students self-selected spaces and phone readings were approximate, this shows association in one school, not causation.[4]

Describes first: groups, centre, n

Test, sign and size; no causal verbs

Qualitative themes explain the pattern

Limitations tied to the claim

REUSABLE TEMPLATE

Your data-analysis plan checklist

Draft one line per item for your own project; adapt the wording for texts, documents or designs.

- [ ] Population, sample and how it was chosen (and who is missing)
- [ ] Cleaning rules: invalid values, missing data, duplicates, decided in advance
- [ ] Descriptive statistics: n, mean or median, spread; or a summary of your texts
- [ ] Analysis: group comparison, correlation (which type and why) or coding approach
- [ ] Uncertainty: instrument precision, sampling error, how you will show it
- [ ] Bias and confounders you can name, and what you did to reduce them
- [ ] Claim check: association or cause? Setting-specific or general?

CHECK YOUR UNDERSTANDING

Knowledge check and next step

Next step: finish your analysis plan and share it with your supervisor

Noor surveyed students already sitting in each space. What kind of sample is this?

A Simple random **B** Stratified **C** Convenience and self-selected **D** Census of all students

What does Spearman's rho = −0.48 tell us in Noor's simulated data?

A Noise causes lower concentration **B** A moderate negative association **C** No relationship **D** 48% of students lost focus

Which is a likely confounder in Noor's study?

A Survey form colour **B** Choosing loud spaces to socialise **C** Number of interviews **D** Using Spearman's rho

Name one way to report uncertainty about phone sound-meter readings.

SOURCES & FURTHER READING

Sources

Illowsky, B., & Dean, S. (2023). 1.2 Data, sampling, and variation in data and sampling. In Introductory Statistics 2e. OpenStax. openstax.org/books/introductory-statistics-2e/pages/1-2-data-sampling-and-variation-in-data-and-sampling

Illowsky, B., & Dean, S. (2023). 1.4 Experimental design and ethics. In Introductory Statistics 2e. OpenStax. openstax.org/books/introductory-statistics-2e/pages/1-4-experimental-design-and-ethics

Illowsky, B., & Dean, S. (2023). 2.7 Measures of the spread of the data. In Introductory Statistics 2e. OpenStax. openstax.org/books/introductory-statistics-2e/pages/2-7-measures-of-the-spread-of-the-data

Illowsky, B., & Dean, S. (2023). 12.3 The regression equation (correlation coefficient). In Introductory Statistics 2e. OpenStax. openstax.org/books/introductory-statistics-2e/pages/12-3-the-regression-equation

Australian Bureau of Statistics. (2023). Correlation and causation. Statistical terms and concepts. abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/correlation-and-causation

Australian Bureau of Statistics. (n.d.). Types of error. Statistical terms and concepts. abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/types-error

Office for National Statistics. (n.d.). Uncertainty and how we measure it for our surveys. ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/uncertaintyandhowwemeasureit

Laerd Statistics. (n.d.). Spearman's rank-order correlation. Lund Research. statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php

Purdue OWL. (n.d.). Analyzing your primary data. Purdue University. owl.purdue.edu/owl/research_and_citation/conducting_research/conducting_primary_research/analyzing_primary_data.html

Braun, V., & Clarke, V. (n.d.). Understanding TA. Thematic Analysis. The University of Auckland. thematicanalysis.net/understanding-ta