EMPOWERING WOMEN
FOR DIGITAL & REMOTE HR CAREERS

MODULE 3 · HR Data, AI Literacy, and Digital Systems · LESSON 3.4

Data quality, limits, and evidence-based reasoning

Learn to recognise when HR data may be unreliable, avoid overinterpreting patterns, and explain clearly what the data can and cannot show.



1.25 training hours · Phase 1 · RemoteHeR

WHAT THIS LESSON COVERS

From unreliable data to result communication

- **Where HR data can become unreliable:** Understand how data quality can change as HR data is collected, entered and processed.
- **Three ways HR data can mislead:** See how small numbers, seasonality, and survivorship can distort conclusions.
- **Correlation and causation:** Recognise why a relationship between two measures does not prove that one caused the other.
- **What the data can and cannot show:** State the evidence and its limits clearly.
- **Presenting an uncomfortable finding:** Communicate an unexpected result without overstating what the data proves.



A number is the end of a process, not the beginning.

HR data is created, entered, changed, combined and reported. At each stage, errors or missing information can affect the result.

Where HR data comes from

From an event to a reported number

1. **Event**
   an application, absence, hire or resignation happens
2. **Entry**
   someone records information in a system
3. **Storage**
   the system stores the information
4. **Processing**
   data is cleaned, combined or calculated
5. **Report**
   a dashboard or report turns it into a number

Where reliability can be lost

The same data can become less reliable at different points

Collection

Information may be missing, incorrect or inconsistent.

Processing

Data may be transferred between systems, combined, categorised, or cleaned.

Interpretation

A dataset can be accurate and still be interpreted incorrectly.

Before you trust a number

Interpretation of HR numbers and indicators

Ask these questions before interpreting an HR number.

- Where did it come from?
- How was it calculated or defined?
- What might be missing?



How can HR data mislead?

These three patterns in HR data deserve extra caution.

Small Numbers

A small change can create a large percentage.

Seasonality

Some measures rise and fall at certain times of the year.

Survivorship

Looking only at people who remain can hide what happened to those who left.



A correct percentage can still give a misleading impression.

If 2 of 4 employees leave, turnover is 50%. The calculation is correct, however four employees are too few to treat the percentage as a stable pattern without further context.



A change over time does not automatically mean a change in performance.

Applications may fall during holiday periods and rise again afterwards. Before calling a drop a problem, compare it with similar periods.

Who is missing from the data can matter as much as who is included.

A survey of current employees tells us about the people who are still employed. It does not directly include employees who already left.





Two things moving together does not prove that one caused the other.

For example, departments with higher training participation may have higher engagement scores and while the data shows an association, it does not, by itself, show that training caused the higher engagement.

Why the manager's conclusion may be too strong

A correlation can have more than one explanation.

Training participation

- influence on engagement
- engaged employees may be more likely to take training
- external factors beside engagement
- coincidental relationship

Engagement

- influenced by training
- influence on employees and indirectly on training
  - external factors beside training
  - coincidental relationship

What would make a causal claim stronger?

A stronger causal claim needs more than a correlation.

1. **Define the question clearly**
   What exactly might be causing what?

2. **Consider alternative explanations**
   What else could produce the same pattern?

3. **Compare appropriate groups or periods**
   Is there a meaningful comparison?

4. **Look for supporting evidence**
   Do other data or analyses support the explanation?

Be precise about what your dataset actually supports.

“Average time-to-hire was 34 days for completed recruitment cases between January and June.” This tells us the average for the cases included in that calculation and the period covered.

What the data cannot show

The same dataset may not explain why the number changed.

Time Span

Why do some roles take longer?

Processes

Did the recruitment process cause the difference?

Missing Data

What happened to cases of excluded data within the set?

Further usage

How does the process continue and what happens to the existing data?

Communicate both sides of the evidence

A useful evidence statement has two parts.

The data shows...

- findings
- population
- period

The data does not show...

- limitations
- unanswered questions
- reasoning

When the finding is uncomfortable

An uncomfortable result still needs an accurate explanation.

What we know

The finding supported by the data.

What we do not know

A limitation that affects the findings.

What we need to investigate

The next question that could help explain the finding.

A difficult result is easier to discuss when the evidence comes first.

State the finding clearly, name its limitation, avoid assigning a cause that the data cannot establish, and suggest the next useful investigation.



If the manager wanted a different answer

You do not need to change the finding to make the conversation easier.

Instead follow these steps:

- Show the evidence first.
- State the limit second.
- Identify the next question third.



NOW IT IS YOUR TURN

What this chart cannot show

Identify what HR data can and cannot support.

Complete an individual worksheet analysing five HR findings. Identify limitations, distinguish correlation from causation, and write an evidence-based response to an uncomfortable finding.

It feeds into

Evidence limitations & interpretation: the assessed artefact's section on what the data can and cannot support.

Read by your tutor

No mark, no score and no ranking. Your tutor reads your work and may give you feedback.

RECAP

Before you move on

- HR data can become unreliable at different points in the data process.
- Small numbers, seasonality and survivorship can make patterns look stronger or clearer than they are.
- Correlation does not establish causation.
- A good finding states both what the data shows and what it cannot show.
- An uncomfortable result should be presented honestly, with its limits and the next question clearly stated.

Complete the activity “What this chart cannot show” and then take the knowledge check.

EMPOWERING WOMEN FOR DIGITAL & REMOTE HR CAREERS

Lesson 3.4 complete

Now you will practise the other half of data literacy: knowing where the evidence stops. You will inspect HR findings, identify what they can support, and state what they cannot show.

mountains of research

ELN MTÜ European Learning Network

egeria desarrollo social

UNIVERSITÄT PADERBORN

Co-funded by the European Union

Erasmus+ 2025-1-PT01-KA220-VET-000365831

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

CC BY-SA 4.0