HUMANITARIAN ACADEMY · HUMANITAIRES JOBS

Mini MBA Programmes and MEAL: design, manage and prove impact

Course handout · Week 7 : Analysing data: useful statistics, pitfalls and visualisation

A programme of exactly 3 months (12 weeks, 144 hours) to design, manage and evaluate humanitarian programmes: needs assessment and severity, results chain and indicators, targeting and cash transfers, planning and budget, MEAL plan and data quality, sampling, analysis, accountability, evaluation, impact measurement, value for money and adaptive management. Corrected numerical exercises, three monthly exams with case studies, and a verifiable diploma from the Humanitarian Academy.

https://humanitairejobs.org/en/academy/mini-mba-programmes-meal/semaine-7/

Free document from the Humanitarian Academy. It may be printed and shared freely for training purposes, without modification.

Week 7 : Analysing data: useful statistics, pitfalls and visualisation

Objectives

- Choose the right descriptive statistics and disaggregate to reveal the inequalities hidden by averages.
- Avoid the classic pitfalls: correlation and causation, Simpson's paradox, misleading denominators.
- Analyse qualitative data rigorously and present results so that they are understood and used.

Descriptive statistics: mean, median, dispersion

The **mean** is sensitive to extreme values; the **median** (the value that splits the sample into two halves) is not. For income, expenditure, debt or waiting times, which are highly skewed distributions, the median better describes the typical household.

Example. Debts of ten households (USD): 0, 0, 20, 30, 40, 50, 60, 80, 100, 1,620. Mean: 2,000 / 10 = 200 USD. Median: (40 + 50) / 2 = 45 USD. Saying "households have an average debt of 200 USD" describes a single household, not the situation.

Dispersion matters as much as the centre: standard deviation, or interquartile range (between the 25th and 75th percentiles) for skewed distributions. Two programmes with an average waiting time of 2 hours are very different if one has waits between 1 h 30 and 2 h 30, and the other between 10 minutes and 7 hours.

For proportion indicators, present the numerator and the denominator: "62% (248 out of 400)". Percentages calculated on very small numbers ("50% of older people", i.e. 2 out of 4) should be avoided or flagged.

> Change in points and in percent: going from 20% to 30% is an increase of 10 percentage points, i.e. a relative increase of 50%. Confusing the two is the most frequent error in reports.

Disaggregating to see inequalities

A satisfactory overall average can hide a group left behind. **Disaggregation** (by sex and age, disability, status, area, household type) is the way to check that the programme reaches everyone.

Example. Overall satisfaction rate: 82%. By group: households without a person with a disability 86%; households with a person with a disability 54%. The programme does not work for an entire group, and the overall figure was masking it.

The **cross-tabulation** is the basic tool: an indicator in rows, a group in columns. To know whether the difference observed between two groups exceeds what sampling chance can produce, a test is used (chi-square test for proportions, t-test or non-parametric test for means) or confidence intervals are compared. Software calculates it; the manager's role is to ask the right question and to interpret.

Beware of **multiple comparisons**: if 20 differences are tested at the 5% threshold, one is expected to be "significant" by pure chance. The main comparisons are therefore defined in advance in the analysis plan, and the others are presented as exploratory.

Correlation, causation and Simpson's paradox

Two variables that move together do not prove that one causes the other. Three alternative explanations always exist: **chance**, a **common factor** (confounder) and **reverse causation**.

Example. Households that took part in agricultural training have better yields. But they are also those with more land and tools, and who were more motivated to sign up: the training may have an effect, but the simple comparison cannot separate it from these initial differences. Measuring an effect requires a counterfactual (week 10).

Simpson's paradox occurs when a trend present in each group reverses when the data are pooled, because of an unequal distribution between groups.

	Centre A	**Centre B**
Severe cases	60 cured out of 100 (60%)	110 cured out of 200 (55%)
Moderate cases	270 cured out of 300 (90%)	170 cured out of 200 (85%)
Total	330 out of 400 (82.5%)	280 out of 400 (70%)

Here, A is better in each group and overall: no paradox. But if B mainly received moderate cases and A mainly severe cases, B could appear better overall while being worse in each group. Rule: compare performance only with **comparable composition**, or by disaggregating by the factor that changes the difficulty.

Analysing qualitative data

Interviews and focus group discussions explain the "why" behind the numbers. Their analysis follows a method, without which one retains only the quotes that confirm what one already thought.

- **Transcription or detailed notes** on the same day, in people's own words.
- **Coding**: codes (themes) are assigned to passages. Start with codes linked to the evaluation questions, and add those that emerge.
- **Matrix**: themes in rows, groups (women, men, young people, displaced, hosts) in columns. This shows which themes recur in which groups.
- **Active search for counter-examples**: cases that contradict the general trend are often the most instructive.

- **Saturation**: data collection stops when new interviews no longer bring new themes.

Qualitative results are not presented as percentages, but frequency can be indicated: "this theme came up in 9 of the 12 focus groups, including all the women's groups". A quote illustrates a finding; it does not prove it.

> Mixed methods: the number says "30% of households do not collect their transfer", the qualitative data say "the cash-out point is 3 hours' walk away and agents often run out of cash". Together, they lead to a decision.

Visualising to make things understood

A chart must answer a question in a few seconds. Practical rules:

- **Horizontal bars** to compare categories, sorted by value; **lines** for trends over time; **maps** for differences between areas; avoid pie charts beyond three categories and 3D effects.
- The bar axis starts at **zero**; otherwise a 2-point difference looks huge.
- A **title that states the message** ("Coverage has fallen in the northern areas since March") rather than a neutral title ("Coverage by area").
- Few colours, one accent colour for what matters, and a marker for the target.
- The source, date and sample size at the bottom of the chart.

For a management dashboard, limit it to 6 to 10 key indicators, each with the current value, the target, the trend and a colour code with defined thresholds. A dashboard with 40 indicators is read by no one.

On a map, show **rates** rather than absolute numbers when areas have different populations: a large city will always have more cases than a village.

From finding to recommendation

A useful analysis follows a three-step progression: **finding** (what the data say), **interpretation** (why, using qualitative data and context) and **recommendation** (what should be done, by whom, when). Recommendations are few, precise and realistic, and each one flows from a finding.

The **level of confidence** must also be stated: strong (several converging sources, sufficient sample), medium, weak (a single source, small sample). Management can decide on medium evidence in an emergency, but it must know that it is doing so.

Finally, results are **presented back** before the final report: a session with the team and partners makes it possible to correct errors of interpretation and to build buy-in for the recommendations. Results are also shared with communities, in an appropriate form (meeting, poster, voice message).

Exercise: analyse a post-distribution monitoring survey

Situation. Post-distribution monitoring of 400 households. Waiting time: mean 2 h 40, median 1 h 20. Overall satisfaction: 78%. Female-headed households (120): 60% satisfied; male-headed households (280): 85.7% satisfied. The average food consumption score rose from 31 to 38; the proportion of households with acceptable consumption rose from 30% to 45%. In the focus group discussions, 8 out of 10 women's groups mention the distribution time (after 3 pm) and the journey home.

Task. 1) What does the gap between the mean and the median waiting time tell you? 2) Express the change in acceptable consumption in points and in percent. 3) What conclusion and what recommendation do you draw on satisfaction?

Answer. A mean much higher than the median indicates that most people wait only a short time, but that one group waits a very long time: you need to identify which sites or which days. Acceptable consumption: +15 percentage points, i.e. +50% in relative terms (this change cannot be attributed to the programme alone without a counterfactual). Satisfaction: check, $0.3 \times 60 + 0.7 \times 85.7 = 78\%$. The 26-point gap between female-headed and male-headed households is large and explained by the focus groups: late distribution times and a dangerous journey home. Recommendation: morning distributions, a priority queue, closer distribution points, then verification at the next monitoring round.

Key takeaways

- Median rather than mean for skewed distributions (income, debt, waiting times).
- A change from 20% to 30% is +10 points, i.e. +50% in relative terms.
- Disaggregation reveals the groups left behind that the average hides.
- Correlation is not causation; compare performance with comparable composition.
- Qualitative analysis: coding, matrix by group, search for counter-examples; a quote illustrates, it does not prove.

Quiz

1. Expenditure of 5 households: 10, 15, 20, 25, 430 USD. Mean and median:
   A. Mean 100, median 20
   B. Mean 20, median 100
   C. Mean 25, median 20
   D. Mean 100, median 25

2. Coverage goes from 40% to 50%. This is:
   A. +10% in relative terms
   B. +10 percentage points, i.e. +25% in relative terms
   C. +50 points
   D. +25 points

3. 40 differences between groups are tested at the 5% threshold. How many "significant" results can be expected by chance alone?
   A. None
   B. About 2
   C. About 20
   D. Exactly 5

4. On a map comparing districts with very different populations, you should show:
   A. The absolute numbers of cases
   B. Rates (for example per 1,000 inhabitants)
   C. Only the names of the districts
   D. NGO budgets

5. A theme comes up in 9 of the 12 focus groups. The correct presentation is:
   A. 75% of the population thinks this
   B. This theme came up in 9 of the 12 groups, specifying which ones
   C. Everyone thinks this
   D. This result is statistically significant

Quiz answers : 1. A · 2. B · 3. B · 4. B · 5. B