HUMANITARIAN ACADEMY · HUMANITAIRES JOBS

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

Course handout · Week 6 : Surveys and sampling: getting the numbers right

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-6/

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

Week 6 : Surveys and sampling: getting the numbers right

Objectives

- Choose a sampling design (simple random, systematic, stratified, cluster) suited to the question and the field.
- Calculate a sample size with margin of error, finite population correction, design effect and non-response.
- Design a reliable questionnaire and correctly interpret a proportion and its confidence interval.

Population, sample and representativeness

A survey seeks to know a **population** (all displaced households in a district) from a **sample**. Results can be generalised only if each unit of the population has a **known, non-zero probability** of being selected: this is probability sampling. A "convenience" sample (households near the road, those present at the distribution) may be large, but it remains biased.

A **sampling frame** is needed first: a list of households, sites or villages, with their population. Without a reliable list, work in stages: select villages or blocks of a camp, then households within them, enumerating them on the spot.

The main designs:

- **Simple random**: random draw from the complete list. The simplest statistically, often impractical in the field.
- **Systematic**: calculate an interval (population / sample size), draw a random starting point, then take one household every "interval". For 2,400 shelters and 240 households to interview, interval = 10; start drawn between 1 and 10, for example 7: shelters 7, 17, 27…
- **Stratified**: the population is divided into groups (displaced and host, urban and rural areas) and a sample is drawn in each group. Essential if reliable results are wanted for each group.
- **Cluster**: first draw clusters (villages), then households in each cluster. Less costly in travel, but less precise: a larger sample is needed.

Calculating sample size

To estimate a proportion, the basic formula is: $\mathbf{n = z^2 \times p \times (1 - p) / e^2}$, where z is 1.96 for a 95% confidence level, p is the expected proportion (0.5 when unknown, because it is the most demanding case) and e the margin of error.

- 5% margin: $1.96^2 \times 0.25 / 0.0025 = 384$ households.
- 7% margin: 196 households.
- 10% margin: 96 households.

Halving the margin multiplies the sample by four. The choice of margin depends on the decision: to track a 10-point change between two surveys, a 10% margin is not enough.

Finite population correction. When the population is small, the required sample decreases: $n = n_0 / (1 + (n_0 - 1) / N)$. For N = 2,000 households: 384 / (1 + 383 / 2,000) = 384 / 1.19 = 323 households. For N = 100,000, the correction is negligible.

Design effect. In cluster sampling, households in the same village resemble each other, which reduces the information each provides. The sample is multiplied by the design effect: **DEFF = 1 + (m − 1) × ICC**, where m is the number of households per cluster and ICC the intra-cluster correlation coefficient. With 10 households per cluster and an ICC of 0.05: DEFF = 1 + 9 × 0.05 = 1.45. Without data, 1.5 to 2 is often used.

Non-response. Divide by the expected response rate: with 10% non-response, n / 0.9.

> Full calculation: 5% margin, p = 0.5, DEFF 1.5, 10% non-response: 384 × 1.5 = 576; 576 / 0.9 = 640 households, for example 64 clusters of 10 households.

Stratification and comparisons

If results are wanted for each stratum with the same precision, the sample size is needed **in each stratum**. For three districts with a 5% margin each, the sample is three times larger than for an overall estimate. This is often the main cause of overly expensive surveys: results are promised by district, by sex of the head of household and by status, without calculating the required sample.

When strata have different sizes and an overall result is wanted, **weighting** is applied: each stratum counts according to its share of the population, not its share of the sample. **Example.** Displaced: 20,000 households, sample 400, 60% food insecure. Host: 80,000 households, sample 400, 30% food insecure. Unweighted average: 45%. Weighted average: 0.2 × 60 + 0.8 × 30 = 36%. Presenting 45% would greatly overstate the situation.

To **compare two surveys** (baseline and endline) or two groups, the sample must be able to detect the expected difference: detecting a change from 40% to 50% with 95% confidence and 80% power requires about 385 households per group in simple random sampling, more with clusters. A small change requires a large sample.

Designing a reliable questionnaire

A questionnaire is built from the indicators and the analysis plan. Essential rules:

- **One idea per question**: "Do you have access to water and soap?" is two questions.
- **Defined recall period**: "In the last 7 days" for food, "in the last 30 days" for expenditure. A period that is too long produces memory errors.
- **Standard modules** for standard indicators (food consumption score, rCSI, Washington Group), unmodified, to remain comparable.
- **Neutral wording**: "What do you think of the distribution?" rather than "Are you satisfied with the distribution?".

- **Order**: simple questions at the beginning, sensitive ones at the end, and never a sensitive question without enumerator training and a referral option.
- **Translation and back-translation** into local languages, then field testing.
- **Length**: beyond 45 minutes, the quality of answers declines; remove what is not essential.

> Informed consent is read to every respondent: purpose of the survey, duration, voluntary nature, no link with aid, confidentiality. A refusal is respected and replaced according to the planned rule, never by the most convenient neighbour.

Interpreting a proportion and its confidence interval

A survey result is an **estimate**, surrounded by uncertainty. The 95% confidence interval of a proportion is approximately: $p \pm 1.96 \times \sqrt{p \times (1 - p) / n}$, multiplied by $\sqrt{DEFF}$ with clusters.

Example. 384 households in simple random sampling, 45% without sufficient access to water. Standard error: $\sqrt{0.45 \times 0.55 / 384} = 0.0254$. Margin: $1.96 \times 0.0254 = 5.0$ points. Interval: 40% to 50%.

Practical consequences:

- If the baseline was 45% (interval 40-50%) and the endline 49% (interval 44-54%), it cannot be claimed that the situation has worsened.
- Results by subgroup have wider intervals, because subgroups are smaller: 45% among 80 female-headed households has a margin of about 11 points.
- Always present the sample size, and the interval for key indicators.

Statistical significance does not mean practical importance: with a very large sample, a 1-point difference can be significant without changing any decision. Conversely, a large difference may not be significant with a small sample: then the data do not allow a conclusion, which is not the same as saying there is no effect.

Organising a field survey

A household survey is prepared in weeks, not days. The typical schedule:

- Week 1: objectives, indicators, analysis plan, sampling plan, authorities' approval.
- Week 2: questionnaire, translation, programming on tablets, tests.
- Week 3: recruitment and training of enumerators (3 to 5 days, with a test and practice), pilot test, corrections.
- Weeks 4 and 5: data collection, with daily quality checks and debriefing meetings.
- Week 6: cleaning, weighting, analysis, validation with the team and partners.
- Week 7: report and feedback, including to communities.

Logistics. Number of enumerators = sample size / (interviews per enumerator per day × days of collection). For 640 households, 4 interviews per enumerator per day, 10 days: 640 / 40 = 16 enumerators, i.e. 4 teams of 4 with one supervisor each. Allow 10 to 20% reserve days for contingencies (rain, security, access).

The **replacement** of absent or refusing households follows a written rule (return once, then replace with the next household in the drawn list); otherwise enumerators replace distant households with nearby ones, which biases the sample.

Exercise: size a baseline survey

Situation. A project wants to know, with a 5% margin and 95% confidence, the proportion of food-insecure households, separately for displaced people (3,000 households) and hosts (40,000 households). Cluster sampling with 12 households per cluster, ICC estimated at 0.06. Expected non-response: 8%.

Task. 1) Calculate the basic sample with the finite population correction for each stratum. 2) Calculate the design effect and the final sample per stratum. 3) How many clusters are needed in total?

Answer. Basic: 384. Displaced: 384 / (1 + 383 / 3,000) = 384 / 1.128 = 340. Hosts: 384 / (1 + 383 / 40,000) = 384 / 1.0096 = 380. Design effect: 1 + 11 × 0.06 = 1.66. Displaced: 340 × 1.66 = 565; 565 / 0.92 = 614. Hosts: 380 × 1.66 = 631; 631 / 0.92 = 686. Total: 1,300 households. Clusters: 614 / 12 = 51.2, i.e. 52 clusters; 686 / 12 = 57.2, i.e. 58 clusters; 110 clusters in total. For an overall result, the two strata will be weighted according to their share of the population (7% and 93%).

Key takeaways

- $n = z^2 \times p \times (1 - p) / e^2$: 384 for a 5% margin at 95% confidence with p = 0.5.
- Correction for a small population: $n = n_0 / (1 + (n_0 - 1) / N)$.
- Cluster design effect: $1 + (m - 1) \times ICC$; then divide by the response rate.
- A result per stratum requires the sample in each stratum; an overall result requires weighting.
- Always present the sample size and the confidence interval; not significant does not mean no effect.

Quiz

1. With p = 0.5, 95% confidence and a 10% margin, the simple random sample size is about:
   A. 38
   B. 96
   C. 196
   D. 384

2. Clusters of 20 households, ICC of 0.04. The design effect is:
   A. 1.04
   B. 1.76
   C. 1.80
   D. 0.80
3. Stratum A: 10,000 households, 50% affected. Stratum B: 30,000 households, 20% affected. The weighted overall proportion is:
   A. 35%
   B. 27.5%
   C. 30%
   D. 25%
4. Baseline: 52% (interval 47-57%). Endline: 55% (interval 50-60%). The most accurate conclusion:
   A. The situation has clearly improved
   B. The data do not allow a change to be concluded
   C. The situation has worsened
   D. The survey is wrong
5. For 2,000 shelters and 200 households to interview with systematic sampling, the interval is:
   A. 5
   B. 10
   C. 20
   D. 200

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