Survey administration

Q 1: What was the survey mode?

a) postal (mailed) survey
b) telephone survey
c) face to face
d) mixed mode

> **Note**: if d): What is the mode of the core questions?

Q 2: Was «computer assisted» interviewing used (CAPI; CATI, etc.)?

a) Yes
b) No

> **Note**: «Yes» should be used only if e.g. skip instructions or consistency checks were part of the computer program.

Q 3: Only for telephone, face to face, and mixed mode surveys: Were self-administered modes used for sensitive questions?

a) Yes
b) No

> ➔ please provide questions for which self-administered questionnaires were used

> **Note**: In face to face surveys these can be self-administered answer sheets or parts of the interview for which the respondent uses the computer to directly answer questions without the help of an interviewer. In telephone interviews new technologies exists, where parts of the interview were conducted by an artificial interviewer (e.g. the interviewer stops the interview and the rest of the interview is automatically conducted by the computer).

Sampling

Q 4: What is the population for which sample should be representative:
age; sex; region; etc.

> **Example**: Non-institutionalized, German-, French- Italian-speaking residents of Switzerland aged 15 and older.

Q 5: Does a sampling frame exist?

a) Yes
b) No

> ➔ please provide description of sampling frame

Definition of sampling frame: list or register of the population elements from which a sample is drawn:

> **Note**: this can be individuals (rare), households (rare); telephone registers; areas (area sampling), municipalities, etc.
> !Probability samples usually need such a list!

Q 6: What is the sampling frame's undercoverage?

These are, for example, homeless people, poor people (no telephone), etc.

Q 7: Was the sample stratified according to one or more criteria?

a) Yes
b) No

➔ Provide variable with stratum identifier

Definition: In a stratified sample separate samples are drawn in each of the exhaustive, non-overlapping subpopulations.

!Stratified sampling covers all sub-populations of the target population (exhaustiveness)!
Example: 2 strata: municipalities over 100'000 inhabitants; municipalities below 100'000 inhabitants.

Q 8: Does sampling uses clusters?

a) Yes
b) No

➔ provide variable with cluster identifier

Definition: Clusters are elements of the subpopulation (or subpopulations) consisting of more than one sampling unit (in alcohol surveys almost always individuals). There is no need to know the sampling frame, but frame can be reconstructed for each cluster, if needed.

Example 1: schools: a list of schools and classes is available, but a list of students per class is not, but can be reconstructed when researchers are in the class

Example 2: households . Addresses of households are known, but not the people living in a household:

Example 3: area sampling; list of areas is used (e.g. in US) but then sampling takes place only in randomly selected areas.

Note: Cluster versus strata: a) cluster sampling is only important if more than 1 person is sampled within a cluster (e.g. more than one person per household; more than one person per area or school or class. b) stratified sampling uses all strata, cluster sampling only a subset of the subpopulation (e.g. only 10 out of 87 areas; only 2'200 schools out of 30'000 schools)

Q 9: What is the primary sampling unit (PSU)?

➔ provide description of PSU

Definition: largest unity of sampling excluding strata, e.g. areas, households, individuals.

Q 10: How many sampling stages?

➔ provide description of sampling stages

Example: sampling of areas(1st stage),sampling of households(2nd stage); sampling of an individual in the household (3rd stage);

Note: when all eligible (see target population; sampling frame) people in a household (e.g. all adults aged 18 or more) were interviewed or approached for interview then this is not a stage, but the household is a cluster.

Nonresponse:

Q 11: How many .....

a) completed interviews

b) partial interviews (interview breakoff, but respondent has answered the interview party)

c) noncontacts of individuals (but it is known that an eligible individual exists)

> **Note**: These may consist of 1) inability to contact person (e.g. target person is in holidays), 2) inability to provide responses (e.g. deaf, mentally ill, does not speak English), 3) refusals

> **Note**: for these cases it is important to know that the unit belongs to the sampling frame; e.g. individual identified by the household roster is eligible for the sample (but in holidays, hospital, etc.)

d) noncontact of household (but valid sampling frame, e.g. valid address, valid telephone number): nobody could be reached

e) non-eligible units: vacant dwellings; vacant units (also seasonally), business units

f) noncontact, no single attempt

> **Note**: This can happen, for example, with commercial pollsters, when a large gross sample was used, but sufficient number of interviews have been already completed (e.g. the client pays for 1'600 Interviews; the gross sample comprises 4'000 addresses, but 1'600 interviews could already be completed by contacting 3'000 households).

g) other

Q 12: What was the maximum number of repeated calls (how often has address, telephone number, etc. been contacted)?

> ➔ If possible, please provide indicator variable of number of contacts per respondent.

> **Note**: This question asked for the number of calls after which an address becomes "a noncontact" (Was this after e.g. 3 or 99 attempts?)

Weights:

Q 13: are pi-weights available?

a) Yes
b) No

> ➔ provide variable with weights

> **Note**: pi-weights are also called design-based weights or probability inclusion weights. These weights are totally independent of non-response. The weights inform about the a priori probability of a person to be included in the sample. The inverse of this weight

stands for the number of people of the target population represented by the corresponding respondent. **An example**: In a simple random sample with a 100% response the pi-weight is n/N for all respondents. In a three stage area sampling (area, households, individuals) the pi-weight must reflect the probability of sampling the area, the probability of sampling the household, and the probability of sampling a person in a household (excluding the non-eligible household members such as minors).

Pi-weights are usually not available if:

- there is no list or register from which the sample is drawn at any of the sampling stages (this is also true if that person in a household is chosen as respondent who first answers the phone call: Note, a complete household roster is needed and a person must be elected randomly from this roster to provide pi-weights.

- randomly selected but refusing households could be replaced by «near by» households (e.g. neighbors)

- ad hoc samples, quota samples

Q 14: Was refusal conversion used?

a) Yes
b) No

➔ Provide % of and indicator variable for converted refusals!

Note: this is important as commercial pollsters in some countries (e.g. The Netherlands) see refusal conversion as unethical. Thus, response rates will be lower in these countries. Refusal conversion means that people initially not willing to be interviewed were «convinced (converted)» by the interviewer to participate.

Q 15: Were initial screening questions used to exclude individuals from the sample?

Example: only people drinking a least once per year were included in the sample. Pi-weights should then still apply to population before screening.

Q16: Who conducted the interviews?

a) commercial pollster
b) federal office
c) students
d)other

Q 17: Was non-response weighting or weighting for sampling frame undercoverage used?

a) non-response weighting
b) weighting for sampling frame undercoverage
c) both
d) none

➔ Please provide weighting variables, and descriptions of variables and description of cells used for weighting (e.g. sex*age with 5 age groups = weighting cells)

Definitions: non-response weighting uses information from the sampling frame only (e.g. 4 of 5 women responded, but only 3 of 5 men, corresponding weights would be 5/4 and 5/3; note

that this needs sampling frame, pi-weights, etc.). Weighting cells can consist of multi-way tables e.g. by sex*age groups.

Weighting for frame undercoverage (usually automatically includes some kind of non-response weighting) uses external information for weighting, e.g. known census data or data from larger scale surveys. Weighting for frame undercoverage is often called poststratification. Cells of the sample (weighted for pi-weights) are compared with known census figures of the same cells (control counts).

Q 18: Description of sampling for non-probability samples:

Definition: Non-probability samples are samples for which pi-weights can not be calculated or non-response can not be determined. **Examples are**: Quota-samples (reviewers receive lists with e.g. sex-age-etc characteristics, for which they have to find respondents, but potential respondents are not selected randomly; clever interviewers find married women with young children near playgrounds or kindergartens); convenience samples (e.g. everybody who responded to the questionnaire Saturday afternoon in the main shopping street); samples for which nonresponents can be replaced by near-by neighbors, etc.

➔ Please provide information about oversampling (note in probability samples this information is reflected in pi-weights), quota used, or any information that can be used to evaluate "representativity" or randomness (note that from my understanding of sampling these are synonyms; GG)

Oversampling means that by design more people were sampled for a subgroup than one would expect from simple probability of that subgroup. **Example**: In Switzerland, some cantons are so small that for a representative Swiss sample only 20-30 individuals would enter the sample by chance. However, cantonal offices may finance a sample of 500 individuals in this canton to get more reliable statistics for their canton.

Miscellaneous:

Q 19: What is the survey year?

Q 20: Drink size information:

What is the average volume of alcohol for an average standard drink in grams (for generic consumption calculation)?

Corresponding measurements (vol-% of beverage, drink size in ml, or directly in grams of pure ethanol) for the various beverage specific drinks (beer, wine, spirits, and others cultural consumed alcohol if they are surveyed)?