SAMPLING

and its relevance for sound data collection
Reasons for survey research

✓ to describe characteristics of certain groups
✓ to make specific predictions (Churchill and Brown 2007: p. 105)
✓ to get an idea about quantities of interest in target population
✓ to report means, totals, proportions for target population (Biemer and Lyberg 2003: p. 49)
Choosing a subset of representative units from the target population (Shao and Zhou 2007: p. 237)

- Important step in data collection process
- Does not get enough attention
- Determines the quality and validity of output
Important definitions

- **Total population**: All elements, sharing some common characteristics, which comprise the universe for the purpose of the research problem (e.g. all organic farmers within a country).

- **Target population**: The group of objects that is identified to participate in the study.

- **Census**: The whole population (e.g. all organic farmers in a country are approached with a questionnaire).

- **Sample**: A subgroup of the whole population is selected for data collection (e.g. every 10th organic farmer on a list of all farmers or only 20 organic farmers in every district are approached).

- **Panel**: A sample of participants who have agreed to provide information at specified intervals over a certain time period.

  according to Malhotra (2012)
Representativeness of a sample

✓ reflecting the target population in all relevant variables (sample as a ‘microcosm’ of the target population)

✓ precondition for generalisation of survey results (Bryman 2008: p. 168)

✓ if not fulfilled for all relevant variables → needs to be communicated (“representative according to age, income etc.”)
The sampling process involves several key steps:

<table>
<thead>
<tr>
<th><strong>target population</strong></th>
<th>The enumeration of objects that possess the necessary information, which needs to be collected (e.g. organic dairy farmers).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>sampling frame</strong></td>
<td>The determination of objects within a target population that will be part of the sampling process (e.g. x farmers in every district).</td>
</tr>
<tr>
<td><strong>sampling technique</strong></td>
<td>Most important decision: Probability or nonprobability sampling (e.g. probability: all farmers have the same chance of being included in the sample; nonprobability: not all farmers have the same chance of being included in the sample, e.g. due to an incomplete address list)?</td>
</tr>
<tr>
<td><strong>sample size</strong></td>
<td>The determination of the sample size is either based on arithmetic means or on proportions (e.g. 100 farmers or 5% of all farmers). Examples of calculations will be presented below.</td>
</tr>
<tr>
<td><strong>sampling process</strong></td>
<td>A specification of how the sampling design (decided upon at a prior stage - shown above) will be implemented (search for complete and up-to-date address lists of organic farmers or approach organic farmers' organisations to receive address lists; nonprobability, if not all farmers are members of an organisation or if one organisation refuses cooperation).</td>
</tr>
<tr>
<td><strong>Validate the sample</strong></td>
<td>Screening of the participants in the data collection phase (e.g. for a study on animal husbandry it makes no sense to include farms without animals; a screening question would then be: &quot;Do you keep animals on your farm?&quot; If no, the interview is finished).</td>
</tr>
</tbody>
</table>
Sampling techniques

- Probability sampling
  - Simple random sampling
  - Systematic sampling
  - Stratified sampling
  - Cluster sampling

- Nonprobability sampling
  - Convenience sampling
  - Judgemental sampling
  - Quota sampling
  - Snowball sampling
## Comparison of probability and nonprobability sampling

<table>
<thead>
<tr>
<th>Probability sampling</th>
<th>Nonprobability sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Sample units are chosen randomly – inclusion of units is</td>
<td>✓ Sample relies on personal judgement</td>
</tr>
<tr>
<td>objective</td>
<td>✓ No calculation of sampling error possible</td>
</tr>
<tr>
<td>✓ Reduced coverage error</td>
<td>✓ Representativeness cannot be ensured</td>
</tr>
<tr>
<td>✓ Sampling error can be calculated</td>
<td>✓ Comparatively low cost</td>
</tr>
<tr>
<td>✓ Usually very expensive</td>
<td>✓ Often used for market research or for preliminary work (e.g. pre-</td>
</tr>
<tr>
<td>✓ Used for aim of representative results, official statistics</td>
<td>test of questionnaires in focus groups)</td>
</tr>
<tr>
<td>✓ With increasing sample size chance increases that sample</td>
<td>✓ More accurate information with very small sample size (&lt; 10 units)</td>
</tr>
<tr>
<td>corresponds in its structure to the basic population</td>
<td></td>
</tr>
</tbody>
</table>

Errors in sampling process

✓ **Sampling error**: An error that occurs, because the selected sample is an imperfect representation of the population.

✓ **Non-sampling error**: An error that can be attributed to sources other than sampling. Non-sampling errors can be random or non-random (e.g. respondents refuse to take part or are not available).

✓ **Random error**: An error that arises from random changes or differences in participants or in measurement situations.
Determination of sample size

✓ Determination of sample size is based on both managerial and financial considerations

✓ No direct relationship between population size and sample size

✓ Typically the larger the sample size, the smaller the sampling error
Important statistical terms

- **Statistic**: Measure of the sample used to describe a certain characteristic (e.g. arithmetic mean, standard deviation)
- **Standard deviation**: Square root of arithmetic mean of individual deviations squared (most common measure of dispersion)
- **Variance**: Standard deviation squared (measure of dispersion)
- **Confidence**: Amount of confidence one wishes to have in estimates
- **Standard error**: Standard deviation of the sampling distribution of the mean
- **Precision**: Possible tolerance of sampling error within a given confidence level
Example: determination of sample size

\[
n = \frac{Z^2 \cdot s^2}{E^2}
\]

n = sample size  
Z = standardised value indicating level of confidence  
s = estimator of population standard deviation  
E = acceptable magnitude of sampling error (precision)

Example: The sample size that is needed to determine the average size of an organic farm in Germany is estimated.

\[
Z = 1.96 \text{ (95\%-level of confidence)}
\]
\[
s = 55 \text{ ha (value derived from a previous study)}
\]
\[
E = 7 \text{ ha (maximum error of ± 7 ha is accepted)}
\]

\[
n = \frac{1.96^2 \cdot 55^2}{7^2} = 237.16 \approx 238
\]

The outcome is a sample size of 238 farms to get a reasonable result on the average size of organic farms in Germany.
Data network for better European organic market information

More information in the Organic Data Network - Manual
Data network for better European organic market information

Additional slides...
Probability sampling I

✓ Every unit of target population has a known and nonzero chance of being included in sample

✓ Sample selection is objective → Sampling error can be estimated

✓ Probability samples have found to be more accurate than nonprobability samples (Shao and Zhou 2007: p. 238)

✓ Probability samples are chosen according to statistical aspects
Probability sampling II

✓ Probability sampling methods are most frequently used for face-to-face, mail, email, and telephone surveys (Lohr, 2010)

✓ Simple random and stratified sampling are basic forms of probability sampling, selection through random procedure, e.g. survey every fifth person on a list of organic dairy farmers

✓ Both are not applicable - necessary to have very good knowledge on the total population (i.e. a complete and up-to-date list of the total population) - otherwise, the sample might have very unusual properties

✓ Stratified sampling and cluster sampling, as well as combinations of both, are most applicable in this context.
Nonprobability sampling I

✓ Sample selection is judgmental, probability of being chosen is unknown
✓ Findings are not projectable to population (Shao and Zhou 2007: p. 238)
✓ Sampling error cannot be assessed (Shao and Zhou 2007: p. 243)
✓ Usually easier and cheaper to conduct than probability sampling methods (Shao and Zhou 2007: p. 238)
Nonprobability sampling II

- Only **quota** sampling might be a relevant option for organic market data collection
- Use of a certain percentage of the target population with particular characteristics of interest (e.g. farm type, farm size, type of product produced)
- Data collector determines the percentage and specifies the number of objects to be included into the sample
- The proportion of objects with a chosen characteristic in the sample should be the same as the proportion in the total population (e.g. if 30% of all dairy farms are located in mountainous areas, then 30% of dairy farms included into the survey should be farms in mountainous areas)
- Good knowledge of the total population is necessary
- Suitable method, if groups with certain characteristics are examined
- Quota plan always has to be set up