Course in consumer research methods

Robert Home, Hanna Stolz, and Claudia Meier
Food Integrity Workshop
Frick, 13-14.03.2018
Program Today

09:00    Tour of FiBL
12:00    Lunch
13:00    Internet catch up time
13:30    Overview of research methods
14:30    Break
15:00    Qualitative Research
17:00    Finish
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:15</td>
<td>General overview</td>
</tr>
<tr>
<td>09:40</td>
<td>Plant science (Hans Jakob Schaerer)</td>
</tr>
<tr>
<td>10:15</td>
<td>SMART (Anja Heidenreich)</td>
</tr>
<tr>
<td>10:30</td>
<td>Soil Sciences (Andreas Fliessbach)</td>
</tr>
<tr>
<td>11:15</td>
<td>Vet sciences (Michael Walkenhorst)</td>
</tr>
<tr>
<td>11:45</td>
<td>International (Toralf Richter)</td>
</tr>
</tbody>
</table>
Overview of research methods

Aims

• Have an overview of what social researchers do
• Intercultural communication

Content

• Research questions
• Qualitative and quantitative methodologies
• Ethical issues
• Research plan
Research questions

Transdisciplinary research

• Involvement of end users in research process
• Including design
• Ensures relevance
• Not common in consumer research

Management questions

Research questions
Management question

Questions that managers ask

Generally very broad

Example: How can we get people to come to our business and spend money?

Usually not much use for research

Reformulate into research questions
Research questions

Central element of research

Usually preceeds methodological framework

Must be accurately and clearly described

Qualitative questions deal with

• What’s going on here?
• How does this work?

Quantitative questions deal with

• How many, or how much?
• Collect data for mathematical analysis
Qualitative research question template

_________ (How or what) is the _________ ("story for" for narrative research; "meaning of" the phenomenon for phenomenology; "theory that explains the process of" for grounded theory; "culture-sharing pattern" for ethnography; "issue" in the "case" for case study) of _________ (central phenomenon) for _________ (participants) at _________ (research site).
Qualitative research question

What is the theory that explains the process of choosing whether to spend money in our business by the general public in the German speaking part of Switzerland?
Quantitative research question

Does ___________ (name the theory) explain the relationship between ___________ (independent variable) and ___________ (dependent variable), controlling for the effects of ___________ (control variable)?
Quantitative research question

Does the theory we developed in our qualitative phase explain the relationship between our offer and people’s choice of whether to patronise our business: controlling for the effects of income, gender, and education?
So. Which methodology?

Research is expensive and time consuming

First question is: Do we need to do primary research?

• Can we get the answers from literature?
  • Secondary research

• Review papers and meta-analyses are common
Secondary research

Faster and cheaper: but only if

• Information is representative
  • E.g. Can we use a study from abroad?
• Information is valid
  • E.g. Can we use this study about a similar product?
  • E.g. Can we use this study from the past?
• Information is reliable
  • Can we trust the quality of the study?

Otherwise primary research needed
Method Choice

Primary research divided into qualitative and quantitative

The choice of the right methodology usually needs a good understanding of the tools available in both qualitative and quantitative paradigms

• What is the problem?
• What do we need to solve the problem?
• How much budget do we have?
• How much time do we have?
Qualitative research uses

Generating hypotheses

Exploring attitudes and needs

Identifying motivations for behaviours (causation, relationships)

Finding out ‘what’s going on here?’

Delivers the ‘important’ items for quantitative research
Qualitative research paradigm

Inductive (generate new theory from data)

Explorative

Small samples (it’s time consuming)

Free from preconceptions (at least we try)

Creative/projective questioning techniques

Not generalizable
Quantitative research uses

Deductive (hypothesis testing)

Identification of correlations

Quantifying: showing who have things in common and how much

Enabling generalisations (inferential statistics, representative samples)

Also providing evidence of causation (experimentally)
Quantitative research paradigm

Deductive

Standardised

Large samples

Limited projective questioning

Hypothesis testing
### Advantages/disadvantages

<table>
<thead>
<tr>
<th>Qualitative</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep understanding of whole concept and relationships</td>
<td>Representative and comparable</td>
</tr>
<tr>
<td>Exploration of new content</td>
<td>Cheap and fast (relatively)</td>
</tr>
<tr>
<td>Individual relevance and values</td>
<td>Less bias</td>
</tr>
<tr>
<td>Not representative</td>
<td>Correlations instead of causality</td>
</tr>
<tr>
<td>Expensive and slow</td>
<td>Dependent on prior knowledge</td>
</tr>
<tr>
<td>Dependent on interviewer skill and participant recruiting</td>
<td>Inflexible (new relevant concepts difficult to integrate)</td>
</tr>
</tbody>
</table>
Ethical issues

Research with human subjects

Helsinki agreement

• All research involving human participants must have been approved by the authors’ Institutional Review Board (IRB) or by equivalent ethics committee(s), and must have been conducted according to the principles expressed in the Declaration of Helsinki. (PLOS One)

But some parts are relevant

• Consent
• Anonymity
• Privacy
• Data ownership
• Recording
Research plan

Describe the issue or problem leading to the research

Review of current knowledge (research gaps)

Indicate deficiencies in the studies.

State the purpose of the research

Outline research questions and hypotheses

Describe what data will be collected

Describe the sample and how the respondents will be recruited
Research plan

Describe how the data will be collected
Describe how the data will be analysed
Describe how the data will be stored
Describe expected results
Implications of results
Describe how ethical issues will be dealt with
Timeline
Budget
Qualitative research

Hanna Stolz
Food Integrity Workshop
Frick, 13.03.2018
Contents

1. Introduction into qualitative research
2. Research objectives
3. Characteristics of qualitative methods
4. Typical research questions
5. Focus groups
6. Qualitative interview
7. Practical part: Research concepts and short presentations
8. Analysis of qualitative data
9. Strengths of qualitative methods
10. Limitations of qualitative methods
11. Discussion/questions
Introduction into qualitative research

- Relevance during the economic boom in the 60ies
  - Market saturation → homogenous products, need for product differentiation
  - Need for methods that are able to analyse various, and often differing needs and expectations of consumers that influence their behaviour
I bought a new dog food, which is supposed to be really good for you.

It contains no meat, just vegetables.

Have you ever considered doing a little market research first?
Introduction into qualitative marketing research

• Qualitative methods are used to reveal motives, needs, perceptions, expectations, subjective/individual opinions

• Suitable to analyse the «soft» (psychological) factors that influence consumer preferences
Research objectives

• In-depth understanding of complex phenomena
• Identify, describe and understand psychological and social connections
• NOT used for measuring them
• “Deep” rather than “broad” data
• Entire picture (spectrum of opinions/arguments) relevant
Characteristics of qualitative methods

- **Open**
  - Structure of interviews is completely open or semi-structured → data collection process is responsive

- **Communicative**
  - adjustment towards interviewees (consumers) and their communication
  - no pre-definition of answers → individual expressions

- **Typifying**
  - analysis of contents/spectra of opinions rather than quantitative characteristics
  - identification of typical opinions/motives, etc.
Characteristics of qualitative methods

• Data collection instruments
  • Open questions
  • Small case numbers
  • Verbal descriptions

• Creative or product-related elements can be included
  • E.g. labels, packages, products
  • Sensory testings
Typical research questions

• How do consumers perceive and evaluate a certain product/brand/packaging/label/etc.?
• What are the various opinions and preferences of a certain product/production method/label?
• Why do consumers prefer one label over another one?
Qualitative research does not study the parts in isolation

Heidi R. Gardner - WordPress.com
Typical research questions

- EU-project ECROPOLIS: The overall objective was to develop an approach to identify target groups for sensory marketing in the participating countries
  - Which senses are important when eating?
  - What kind of pictures comes into consumers mind when consuming organic products?
  - Do consumers perceive sensory differences to conventional food when eating organic food?
  - Do consumers expect organic products to taste similarly to conventional products, or differently?
  - In which occasions do consumers prefer organic or conventional products because of their sensory properties?
  - Do they remember situations where their buying decision was influenced by sensory information and how?
Data collection methods

• Focus groups

• Qualitative interview
Focus groups

• Most commonly used qualitative method
• Interviews that are conducted with a group of consumers (not with individuals)
  • group size depends on the topic
  • usually 8-12 consumers per group (delicate topics require smaller group sizes)
• Several groups (at least 4) → saturation of discussion contents
• Focussed on a pre-defined topic
• Lead by a facilitator
• Usually take about 1-2 hours
Focus groups

• Communicative group interaction
  → identification of range of opinions and arguments
  → consumers feel stronger in groups and thus are more free to express their personal opinions
  → interviewer bias is relatively low
• Consumer attitudes, perceptions, preferences, motives and habits are object of the discussions
• Time and cost effective method
Focus groups

Focus groups are used for…

• structuring, predicting, idea generation, screenings, identification of causes, and explorations of new topics

• when group dynamic effects are relevant for the research context or when an group opinion forming process is requested

• when lots of material is used and the spontaneous and emotional opinions shall be collected

• when a range of (controversial) opinions/reactions shall be revealed

• for improving concepts since participants inspire each other and improve ideas
Focus groups

Conducting focus groups require

• Sampling and recruitment of participants
• Location
• Interview guidelines
• Facilitator
Focus groups

Sampling and recruitment

• Define target group
• Make sure that participants are representative (quota for age, gender, etc.), recruitment questionnaire (quota and screening questions)
• Participants should not know each other
• Recruitment through marketing agency
Focus groups

Location

• Studio/room in mostly neutral and silent environment, centrally (easy to access)
• round table
• audio or video tape
Focus groups

Interview guidelines

• Introduction
• Sequences with leading questions (open questions)
• Within leading questions, more specific questions for stimulation the discussion or for leading towards relevant topics)
• Ice breaker question (collect opinion of every participant in a first round)
• Time frame
Focus groups

Facilitator

- Leads discussion
- Neutral!
- Not taking part in the discussion!
- Stimulate discussion, calm down strong opinion leaders and encourage silent participants to express their opinions
- Makes sure that topics addressed in the discussion are relevant and interrupt carefully, if necessary
Focus groups

Summary

- Most relevant marketing research method
- Flexible tool to analyse the range of attitudes/motives etc.
- Relatively small interviewer bias
Qualitative interview

When using qualitative interviews? If…

• Individual level is relevant
• Background and norms of individuals are relevant
• Processes as a whole are analysed (e.g. buying decision process)
• Difficult to recruit participants (experts)
• Specific techniques shall be used, e.g. laddering technique
• Research question or question technique requires concentration
• In-depth interview
Qualitative interview

- Qualitative interview or (semi-structured interview) usually based on pre-defined interview guidelines
- Pre-defined target group (recruitment guidelines based on quota and screening questions)
- No group interaction
- Interviewer requirements partly same as of focus group facilitator
- Audio tape
- Try to reduce interviewer bias
Interview techniques of qualitative methods

Laddering interview

• technique from psychology to encourage self-analysis of behaviour and motivations
• to gather a more complete list of “consequences” and climb towards the hard-to-reach “values” (to predict behaviour).
• Example:
  • 1. “Why did you choose this product/service?” – to establish the important attributes
  • 2. “Why is it good/bad that…?” – to establish the consequences of each attribute
  • 3. “Why is this important to you?” or “How does this relate to your core values?” – to establish the values of the respondent which are affected by each consequence
Interview techniques of qualitative methods

Indirect question

• E.g. What do you think your neighbour would think of product x?
Practical part

- Group work: 2-3 persons
- Draft concept for a qualitative survey
  - Research objective
  - Research questions
  - Focus group/interview guidelines (short)
- Short presentations

- Duration: 30 min + 30 min
Analysis of qualitative data

• Explicatory methods

• Qualitative content analysis
Analysis of qualitative data

- Reproduction of results is more difficult than within quantitative methods
- Constant and standardised/systematic analysis process necessary
- Objective analysis process (several persons involved in the analysis/repeat analysis)
- Make data analysis and interpretation transparent
Analysis of qualitative data

Qualitative content analysis

- Evaluation method of hermeneutics
  - to understand human behaviour
  - methodology for interpreting verbal transcripts
  - constructive and reproductive process
  - systematic structuring

- Used to objectively structure complex qualitative data material

- Comparing material from different interviews/focus groups or even from different studies

- Software: e.g. Nvivo, ATLAS.ti, MAXQDA
Analysis of qualitative data

Procedure

1. Transcription of focus group discussions/interviews using standard orthography
2. Elaboration of a coding system including a list of contents to be analysed
3. Assignment of certain code to each issue
4. Systematic coding of statements of individuals
5. Structuring statements according to code system
6. Conduction of content analysis, involving classifying, structuring and paraphrasing from the individual arguments to a general sense and relevance
Strengths of qualitative methods

• Identification of unnoticed connections and exploration of new themes
• In-depth analysis of perception, attitudes and other psychological constructs
• Get the “whole story”
Limitations

• Limited representativity, generalisability and quantifiability of results
• Potential risk of subjectivity in data interpretation
• Potential lack of traceability in data interpretation
• Complex or spontaneous behaviour requires other methods
Excursion

Ethnography in qualitative marketing research

• To reveal/analyse complex or unconscious or affective (spontaneous) decision making processes
• Observing and interviewing consumers in their real environment
  • At home, e.g. when preparing or consuming food
  • In shops
What can be done with the results

• Marketing recommendations
  • Product
  • Promotion/communication
• Building a theory
• Serve as basis for quantitative research
• Recommendations for decision-makers
• Dissemination of results
Methods for the collection and analysis of quantitative data: quantitative survey research

Claudia Meier (claudia.meier@fibl.org)

Food Integrity Training

FiBL, 13./14.03.2018
Quantitative survey research

Define: ...

1. Research questions
2. Hypotheses
3. Target population
4. Sampling
5. Methods for data collection
6. The data collection tool: the questionnaire
7. Methods for data analysis
Research questions
Research questions

What are consumers’ preferences with respect to a product or service?

Examples from our research:

- Is the rural population in southern India willing to pay more for a millet variety with a higher iron content? What are the determinants of WTP?

- Are Europeans willing to pay more for vegetable varieties, which are bred by the farmers themselves, so-called farmers’ vegetable varieties? What are the determinants of WTP?

- Is there a significant difference in the consumer acceptance of meet products from 95% and 100% organic feeding? What are the determinants of consumer acceptance?
Hypotheses
Hypotheses

**Research question:** Are Europeans willing to pay more for farmers’ vegetable varieties? What are the determinants of WTP?

- **Null hypothesis:** Europeans are not willing to pay more for farmers’ vegetable varieties \((WTP \leq 0)\)
- **Alternative hypothesis:** Europeans are willing to pay more for farmers’ vegetable varieties \((WTP > 0)\)
Target population
Target population

Research question: Are Europeans willing to pay more for farmers’ vegetable varieties? What are the determinants of WTP?

- **Target population** = all the people or objects about which we want to know more → results will only be valid for this group!

- **Describe your target population:**
  - Geographically: living in Europe (Switzerland, France, Italy, Spain)
  - Demographically: between 18 and 79 years old
  - Object-related: (jointly) responsible for food purchase; consume vegetables
  - To avoid conflicts of interests: not working in marketing, market research, food production/processing, wholesale, retail

→ **Draw a sample from the target population that is representative for the target group!**
Side remark: When can we say that a sample is representative of the target group?

• A sample is representative for the target group if it is composed in such a way that it has an identical structure to the target group in terms of essential characteristics (e.g. Age, Gender, place of residence, occupation).

• This will most likely be the case if the sampling procedure is applied correctly and a sufficiently large sample is examined. A further condition is that as many people as possible take part in the study, i.e., that a high response rate is achieved.
Short exercise

- **Research question:** What are the consumer preferences for wine in different neighbourhoods of the city of Zurich? I.e. which types of wine and wine brands are preferably consumed in different neighbourhoods?

- **Target population:** Define!

- **Tought experiment:** Every neighbourhood in Zurich has its own recycling station for glass bottles/jars etc. Could we elicit a neighbourhood’s wine preferences by examining the wine bottles which are taken to the neighbourhood’s recycling station from April 1st to April 30st?

- **In other words:** Is the population who takes the empty wine bottles to a neighbourhoods recycling station representative for the wine consuming population of that same neighbourhood?
Sampling
Sampling

Sampling methods

Random sampling

- Special forms of random sampling:
  - Stratified random sampling
  - Cluster sampling
  - Random-Route sampling

Quota sampling

- Mixed form:
  - Random-Quota sampling

- Special forms of quota sampling
  - Snowball Sampling
Sampling
Sampling methods – random sampling

Rule: «Each element of the target population must have the same chance of being included in the sample.”

Procedure:
1. Basis: Complete and current address file for the target population
2. Number all the addresses
3. Define gross sample size with some buffer for invalid addresses
4. Draw the gross sample from the target population using a random number generator

Problem: if response rate is low → people who do not want or cannot participate cannot be replaced! Multiple contact attempts!
Sampling

Sampling methods – special forms of random sampling

- **Stratified random sampling**: divide the members of the target population into homogeneous subgroups (strata) before sampling (e.g. according to their country of origin). → if target population is very heterogeneous in terms of consumer preferences

- **Cluster sampling**: divide the members of the target population into geographical clusters (e.g. districts, neighbourhoods), draw a random sample from those clusters, either survey all people of the selected clusters (one-stage sampling) or draw a sample of people within each cluster (two-stage sampling). → if geographic concentration is important
Sampling
Sampling methods – quota sampling

Procedure

1. Define 2 to 4 quota characteristics for the selection of respondents (e.g. Age, Gender, place of residence).
   - The sample has an identical structure to the target group in terms of …
   - depending on the study other quota characteristics can be relevant (e.g. customer/non-customer, organic consumer/non-organic consumer or demographics like occupation, education, household size, income etc.)

2. Determine the percentages of the individual quota characteristics and their attributes, i.e. the structure within the target population with respect to the quota characteristics and their attributes.

3. Determine the number of interviews per quota cell for a given sample size.

Problem: if quota chosen turn out not to be relevant for the study
### Sampling
#### Sampling methods – quota sampling – example

**Target population:** 20 to 49 years old, living in the german speaking part of Switzerland.

**Sample size:** \( n = 300 \)

1. Relevant quota: Age and Gender
2. Percentages for every combination of the quota characteristics’ attributes (source: e.g. Federal Statistical Office)

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 to 29 years</td>
<td>Male</td>
<td>15.2%</td>
<td>14.8%</td>
<td>30.0%</td>
</tr>
<tr>
<td>30 to 39 years</td>
<td>Female</td>
<td>16.2%</td>
<td>15.9%</td>
<td>32.1%</td>
</tr>
<tr>
<td>40 to 49 years</td>
<td>Male</td>
<td>19.2%</td>
<td>18.7%</td>
<td>37.9%</td>
</tr>
<tr>
<td>Total</td>
<td>Female</td>
<td>50.6%</td>
<td>49.4%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Sampling
Sampling methods – quota sampling – example

3. Number of interviews per quota cell (n = 300)

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 to 29 years</td>
<td>Male</td>
<td>46 (15.2%)</td>
<td>44 (14.8%)</td>
<td>90 (30.0%)</td>
</tr>
<tr>
<td>30 to 39 years</td>
<td>Male</td>
<td>49 (16.2%)</td>
<td>47 (15.9%)</td>
<td>96 (32.1%)</td>
</tr>
<tr>
<td>40 to 49 years</td>
<td>Male</td>
<td>58 (19.2%)</td>
<td>56 (18.7%)</td>
<td>114 (37.9%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>152 (50.6%)</td>
<td>148 (49.4%)</td>
<td>300 (100%)</td>
</tr>
</tbody>
</table>

→ This sampling method has become well established in everyday practice.
 Sampling
 Sampling methods – special forms of quota sampling

• **Snowball sampling:** used if target group is very small, e.g. people who only consume organic food. Start with a small address file and enlarge it through the contacts of your respondents.
Sampling
Sample size

Target population

true parameter \( \mu \)
(= population mean)

Inference

Random selection

Sample

statistic \( \bar{x} \)
(= sample mean)
Sampling
Sample size – formulas to calculate CI’s – mean values

\[ \pm e = \pm t \times \frac{s}{\sqrt{n}} \]

Legend:
- \( e \): sampling error
- \( t \): critical value (for a CI of 95% \( t = 1.96 \))
- \( s \): standard deviation (= measure of dispersion/variation of \( x_i \))
- \( \bar{x} \): mean
- \( n \): sample size

Definition: The sampling error is the random deviation of a sample result from the «true» value.

\[ s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \]

Example: The average age in a sample of \( n = 508 \) is 46, with a standard deviation of 15. With a probability of 95% the «true» value lies between 45 and 47 (± 1).
Sampling
Sample size – formulas to calculate CI’s – percentages

\[ \pm e = \pm t \times \sqrt{\frac{p \times q}{n}} \]

**Legend:**
\( \pm e \) = sampling error in %
\( t \) = critical value (for a CI of 95% \( t = 1.96 \))
\( p \) = share of respondents with a particular characteristic in %
\( q = 100\% - p \)
\( n \) = sample size

**Definition:** The sampling error is the random deviation of a sample result from the «true» value.

Example: Share of women in a sample of \( n = 300 \) is 49%. With a probability of 95% the «true» value lies between 43% and 55% (± 6%).
Sampling
Sample size – formulas to calculate CI’s – percentages

Note:

• The size of the sampling error depends on the sample size n and the percentage shares p and q. Generally: the larger the sample size and the larger the difference between the share of people with (p) and without (q) a certain characteristic, the smaller the sampling error.

• The product p*q is largest if p equals 50%. Important: If p is not known in advance, which is usually the case, p is always assumed to be 50%.
Sampling
Sample size – formulas to calculate CI’s

Note:

• Definition of sampling error: The sampling error is the random deviation of a sample result from the «true» value. It is denoted by e or \( \sigma \).

• Confidence interval, \( \pm e \) (sampling error, random error, standard error, statistical error), error margin, range of fluctuation can all be used synonymous.

• In theory the formulas to calculate the sampling error are only valid in the case of simple random sampling. However, in practice they are also used in the case of other sampling techniques.

• The larger the sample size the smaller the sampling error.

• The quality of a sample is not proportional to its size, but increases with the square root.
Sampling
Sample size – formulas to calculate CI’s – mean values

Sample size?

Formula for mean values:

\[ \pm e = \pm t * \frac{s}{\sqrt{n}} \]

\[ n = \frac{t^2 s^2}{e^2} \]

Formula for percentages:

\[ \pm e = \pm t * \sqrt{\frac{p * q}{n}} \]

\[ n = \frac{t^2 p * (100-p)}{e^2} \]
Sampling
Sample size – significant differences

Formula for mean values:

\[ \pm e = \pm t \times \frac{s}{\sqrt{n}} \]
Sampling
Sample size

Which groups do you want to compare?
«young vs. old»/ «female vs. male»/ «respondents who consume organic on a regular basis vs. respondents who consume organic on a irregular basis» etc.

Are there enough respondents in each group?
• At least 30 respondents per group

Do we have to boost certain groups?
## Sampling

### Sample size

#### Sample size of n = 300:

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 to 29 years</td>
<td>Male</td>
<td>46 (15.2%)</td>
<td>44 (14.8%)</td>
<td>90 (30.0%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>44 (14.8%)</td>
<td>46 (15.2%)</td>
<td></td>
</tr>
<tr>
<td>30 to 39 years</td>
<td>Male</td>
<td>49 (16.2%)</td>
<td>47 (15.9%)</td>
<td>96 (32.1%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>47 (15.9%)</td>
<td>49 (16.2%)</td>
<td></td>
</tr>
<tr>
<td>40 to 49 years</td>
<td>Male</td>
<td>58 (19.2%)</td>
<td>56 (18.7%)</td>
<td>114 (37.9%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>56 (18.7%)</td>
<td>58 (19.2%)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Male</td>
<td>152 (50.6%)</td>
<td>148 (49.4%)</td>
<td>300 (100%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>148 (49.4%)</td>
<td>152 (50.6%)</td>
<td></td>
</tr>
</tbody>
</table>

#### Sample size of n = 200:

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 to 29 years</td>
<td>Male</td>
<td>30 (15.2%)</td>
<td>30 (14.8%)</td>
<td>60 (30.0%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>30 (14.8%)</td>
<td>30 (15.2%)</td>
<td></td>
</tr>
<tr>
<td>30 to 39 years</td>
<td>Male</td>
<td>33 (16.2%)</td>
<td>32 (15.9%)</td>
<td>64 (32.1%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>32 (15.9%)</td>
<td>33 (16.2%)</td>
<td></td>
</tr>
<tr>
<td>40 to 49 years</td>
<td>Male</td>
<td>38 (19.2%)</td>
<td>37 (18.7%)</td>
<td>76 (37.9%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>37 (18.7%)</td>
<td>38 (19.2%)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Male</td>
<td>101 (50.6%)</td>
<td>99 (49.4%)</td>
<td>200 (100%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>99 (49.4%)</td>
<td>101 (50.6%)</td>
<td></td>
</tr>
</tbody>
</table>
Methods for the collection of quantitative data
Methods for the collection of quantitative data

Quantitative survey methods

By phone: CATI

Online: CAWI

Offline: Face-to-face
  «in home»
  «in hall»

Offline: On paper
Methods for the collection of quantitative data
Evaluation criteria

1. Representativeness:
   • «How well does the sample represent the target population? Is the probability to be sampled equal for every individual in the t.p.?»

2. Validity:
   • «How well do the results correspond to the truth? Do we measure what we intend to measure? Is there a systematic bias?»

3. Reliability:
   • «How good is the quality of our data? How precise is our data? How much noise do we have in our data?»

4. Visual/ audio/ sensory aids and stimuli
   • «Do we need to show pictures/ videos or other display material? Do we want to make people taste something?»
Methods for the collection of quantitative data
Evaluation criteria

5. Survey length and complexity
• “How many questions can we ask? How complex can the questionnaire be (filtering)?”

6. Costs per Interview
• “How much does the data collection cost?”

7. Timing
• “How long does the data collection take?”
## By phone CATI

### Criteria Evaluation

#### 1. Representativeness

- **Coverage**: List with all phone numbers from all participants as a basis for selection? 78% registered fixed line numbers; 98% coverage using fixed line RDD and mobile RDD → good coverage (using RDD)

- **Equal probability to be chosen**: Random selection of household and of person within household → random sample, no geographical clusters

- **Response rate**: Realised int./attempted int.: multiple contacts possible (arrange an appointment); active motivation by interviewers → high response rate

#### 2. Validity

- **Understanding of questions**: Possibility for researcher to listen in on several calls (during pretest and field) and verify understanding; standardized assistance by interviewers to assure correct understanding

- **Interviewer bias**: Interviewer is invisible

- **Tendency to give socially desired answers**: Respondents are anonymous → social pressure is low

#### 3. Reliability

- **Possibility for large samples**: n = 500 to 100,000 possible (compensation for noise in the data)

- **Training and Supervision**: Interviewers have to be trained and can be monitored

- **Computer assisted interview**: Answers can be checked for plausibility

---

0 = bad; + = medium; ++ = good; +++ = very good
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Visual aids: only when previously sent by post&lt;br&gt;Acoustic/ audio aids: possible&lt;br&gt;Tasting: not possible</td>
</tr>
<tr>
<td>5</td>
<td>Maximal survey length: 30 minutes; ideally 20 minutes&lt;br&gt;Complexity: can be high due to computer assisted programming of questionnaire</td>
</tr>
<tr>
<td>6</td>
<td>Infrastructure needed: rooms, technical equipment&lt;br&gt;Staff needed: interviewers</td>
</tr>
<tr>
<td>7</td>
<td>Timing: Depends on lab capacity</td>
</tr>
</tbody>
</table>
By phone
CATI

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 1 Representativeness | **Coverage** (list with all phone numbers from all participants as a basis for selection?): 78% registered fixed line numbers; 98% coverage using fixed line RDD and mobile RDD → good coverage (using RDD)  
**Equal probability to be chosen**: random selection of household and of person within household → random sample, no geographical clusters  
**Response rate** (realised int./attempted int.): multiple contacts possible (arrange an appointment); active motivation by interviewers → high response rate |
| 2 Validity  | **Understanding of questions**: possibility for researcher to listen in on several calls (during pretest and field) and verify understanding; standardized assistance by interviewers to assure correct understanding  
**Interviewer bias**: interviewer is invisible  
**Tendency to give socially desired answers**: respondents are anonymous → social pressure is low |
| 3 Reliability | **Possibility for large samples**: n = 500 to 100’000 possible (compensation for noise in the data)  
**Training and Supervision**: Interviewers have to be trained and can be monitored  
**Computer assisted interview**: answers can be checked for plausibility |
### Criteria Evaluation

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 4        | Visual/ audio/ sensory aids and stimuli | **Visual aids:** only when previously sent by post  
**Acoustic/ audio aids:** possible  
**Tasting:** not possible |
| 5        | Survey length and complexity | **Maximal survey length:** 30 minutes; ideally 20 minutes  
**Complexity:** can be high due to computer assisted programming of questionnaire |
| 6        | Costs per interview vs. information gained | **Infrastructure needed:** rooms, technical equipment  
**Staff needed:** interviewers |
| 7        | Timing | Depends on lab capacity |
### Online CAWI

0 = bad; + = medium; ++ = good; +++ = very good

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 1 Representativeness | **Coverage** only internet users! → for 15 to 64 year olds almost full coverage  
**Equal probability to be chosen**: depends on the sourcing of the respondents.  
Use actively/ offline (by phone) recruited Access Panel (no bias towards heavy internet users)  
**Response rate** (realised int./attempted int.): depends on how panelists have been recruited, how the panel is managed, the incentive and study topic |
| 2 Validity        | **Understanding of questions**: possibility for researcher to do a pretest and verify understanding; possibility to provide additional information below the question text or using pop-up windows  
**Interviewer bias**: no interviewer  
**Tendency to give socially desired answers**: respondents are anonymous → social pressure is low |
| 3 Reliability     | **Possibility for large samples**: n = 500 to 100'000 possible (compensation for noise in the data)  
**Training and Supervision**: not necessary, therefore the work of the interviewers as source for mistakes is eliminated  
**Computer assisted interview**: answers can be checked for plausibility → still: «Durchklicker/speeding» is a problem  
There is no guarantee about «who» answers the questionnaire → personalized questionnaire links! |
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td><strong>Visual/ audio/ sensory aids and stimuli</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Visual aids:</strong> good</td>
</tr>
<tr>
<td></td>
<td><strong>Acoustic/ audio aids:</strong> good</td>
</tr>
<tr>
<td></td>
<td><strong>Tasting:</strong> not possible</td>
</tr>
<tr>
<td></td>
<td>→ Possibility for gamification</td>
</tr>
<tr>
<td>5</td>
<td><strong>Survey length and complexity</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Maximal survey length:</strong> 20 minutes</td>
</tr>
<tr>
<td></td>
<td><strong>Complexity:</strong> can be high due to computer assisted programming of questionnaire</td>
</tr>
<tr>
<td>6</td>
<td><strong>Costs per interview vs. information gained</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Infrastructure needed:</strong> none</td>
</tr>
<tr>
<td></td>
<td><strong>Staff needed:</strong> none</td>
</tr>
<tr>
<td>7</td>
<td><strong>Timing</strong></td>
</tr>
<tr>
<td></td>
<td>Fast (10 days minimum for quality of results)</td>
</tr>
</tbody>
</table>

0 = bad; + = medium; ++ = good; +++ = very good
## Criteria Evaluation

**1 Representativeness**

| ++ | Coverage: only internet users! → for 15 to 64 year olds almost full coverage
|    | Equal probability to be chosen: depends on the sourcing of the respondents. Use actively/ offline (by phone) recruited Access Panel (no bias towards heavy internet users)
|    | Response rate (realised int./attempted int.): depends on how panelists have been recruited, how the panel is managed, the incentive and study topic |

**2 Validity**

| +++ | Understanding of questions: possibility for researcher to do a pretest and verify understanding; possibility to provide additional information below the question text or using pop-up windows
|    | Interviewer bias: no interviewer
|    | Tendency to give socially desired answers: respondents are anonymous → social pressure is low |

**3 Reliability**

| ++ | Possibility for large samples: \( n = 500 \) to 100,000 possible (compensation for noise in the data)
|    | Training and Supervision: not necessary, therefore the work of the interviewers as source for mistakes is eliminated
|    | Computer assisted interview: answers can be checked for plausibility → still: «Durchklicker/speeding» is a problem
|    | There is no guarantee about «who» answers the questionnaire → personalized questionnaire links!

---

0 = bad; + = medium; ++ = good; +++ = very good
# Online CAWI

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 4        | Visual aids: good  
            **Acoustic/ audio aids:** good  
            **Tasting:** not possible  
            → Possibility for gamification |
| 5        | **Maximal survey length:** 20 minutes  
            **Complexity:** can be high due to computer assisted programming of questionnaire |
| 6        | **Infrastructure needed:** none  
            **Staff needed:** none |
| 7        | Fast (10 days minimum for quality of results) |

0 = bad; + = medium; ++ = good; +++ = very good
### Offline
 **Face-to-Face «in home»** 0 = bad; + = medium; ++ = good; +++ = very good

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 1 Representativeness | **Coverage** 100%  
**Equal probability to be chosen**: random selection of household and of person within household  
but: geographical clusters  
**Response rate** (realised int./attempted int.): multiple contacts possible but usually not economically viable |
| 2 Validity    | **Understanding of questions**: pretest to verify understanding; standardized assistance by interviewers to assure correct understanding  
**Interviewer bias**: interviewer is visible, therefore high risk of interviewer bias  
**Tendency to give socially desired answers**: respondents are not anonymous  
→ social pressure is high |
| 3 Reliability | **Possibility for large samples**: No, usually < 500 (no compensation for noise in the data)  
**Training and Supervision**: Interviewers have to be trained, but monitoring is difficult (does interviewer choose respondents correctly, is the interview conducted correctly?)  
**Computer assisted interview**: answers can be checked for plausibility |
**Offline Face-to-Face «in home»**

0 = bad; + = medium; ++ = good; +++ = very good

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Visual/ audio/ sensory aids and stimuli</td>
</tr>
<tr>
<td></td>
<td><strong>Visual aids</strong>: possible</td>
</tr>
<tr>
<td></td>
<td><strong>Acoustic/ audio aids</strong>: possible</td>
</tr>
<tr>
<td></td>
<td><strong>Tasting</strong>: possible (in theory)</td>
</tr>
<tr>
<td>5</td>
<td>Survey length and complexity</td>
</tr>
<tr>
<td></td>
<td><strong>Maximal survey length</strong>: up to 1 hour</td>
</tr>
<tr>
<td></td>
<td><strong>Complexity</strong>: can be high due to computer assisted programming of questionnaire</td>
</tr>
<tr>
<td>6</td>
<td>Costs per interview vs. information gained</td>
</tr>
<tr>
<td></td>
<td><strong>Infrastructure needed</strong>: none</td>
</tr>
<tr>
<td></td>
<td><strong>Staff needed</strong>: interviewers</td>
</tr>
<tr>
<td></td>
<td>Very high costs!</td>
</tr>
<tr>
<td>7</td>
<td>Timing</td>
</tr>
<tr>
<td></td>
<td>Very time intensive</td>
</tr>
</tbody>
</table>
**Offline**
**Face-to-Face «in home»** 0 = bad; + = medium; ++ = good; +++ = very good

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 1 Representativeness | **Coverage** 100%  
**Equal probability to be chosen**: random selection of household and of person within household → but: geographical clusters  
**Response rate** (realised int./attempted int.): multiple contacts possible but usually not economically viable |
| 2 Validity     | **Understanding of questions**: pretest to verify understanding; standardized assistance by interviewers to assure correct understanding  
**Interviewer bias**: interviewer is visible, therefore high risk of interviewer bias  
**Tendency to give socially desired answers**: respondents are not anonymous → social pressure is high |
| 3 Reliability  | **Possibility for large samples**: No, usually < 500 (no compensation for noise in the data)  
**Training and Supervision**: Interviewers have to be trained, but monitoring is difficult (does interviewer choose respondents correctly, is the interview conducted correctly?)  
**Computer assisted interview**: answers can be checked for plausibility |
### Offline
**Face-to-Face «in home»**

0 = bad; + = medium; ++ = good; +++ = very good

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 4 | Visual/ audio/ sensory aids and stimuli  
*** | Visual aids: possible  
Acoustic/ audio aids: possible  
Tasting: possible (in theory)  
*** |
| 5 | Survey length and complexity  
*** | Maximal survey length: up to 1 hour  
Complexity: can be high due to computer assisted programming of questionnaire  
*** |
| 6 | Costs per interview vs. information gained  
0 | Infrastructure needed: none  
Staff needed: interviewers  
Very high costs!  
0 |
| 7 | Timing  
0 | Very time intensive |
## Offline
Face-to-Face «in hall»

### Criteria Evaluation

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| **Representativeness** | **Coverage 100%**  
*Equal probability to be chosen:* geographical clustering; bias towards mobile people  
*Response rate* (realised int./attempted int.): multiple contacts not possible; people are busy |
| **Validity**      | **Understanding of questions:** pretest to verify understanding; standardized assistance by interviewers to assure correct understanding; in a teststudio: researcher can listen in on interviews  
*Interviewer bias:* interviewer is visible, therefore high risk of interviewer bias  
*Tendency to give socially desired answers:* respondents are not anonymous → social pressure is high |
| **Reliability**   | **Possibility for large samples:** No, usually < 500 (no compensation for noise in the data)  
*Training and Supervision:* Interviewers have to be trained, monitoring is better than with «in home» interviews; monitoring is very good in the teststudio (does interviewer choose respondents correctly, is the interview conducted correctly?)  
*Computer assisted interview:* answers can be checked for plausibility |
### Offline
#### Face-to-Face «in hall»

0 = bad; + = medium; ++ = good; +++ = very good

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 4        | **Visual aids**: possible  
          | **Acoustic/ audio aids**: possible  
          | **Tasting**: possible (particularly in teststudio) |
| 5        | **Maximal survey length**: 20 minutes (if respondents are previously length can be increased)  
          | **Complexity**: can be high due to computer assisted programming of questionnaire |
| 6        | **Infrastructure needed**: if teststudio  
          | **Staff needed**: interviewers |
| 7        | Very time intensive |
# Offline

## Face-to-Face «in hall»

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| **1** Representativeness | **Coverage** 100%  
*Equal probability to be chosen:* geographical clustering; bias towards mobile people  
*Response rate* (realised int./attempted int.): multiple contacts not possible; people are busy |
| ++ |  |

| **2** Validity | **Understanding of questions:** pretest to verify understanding; standardized assistance by interviewers to assure correct understanding; in a teststudio: researcher can listen in on interviews  
*Interviewer bias:* interviewer is visible, therefore high risk of interviewer bias  
*Tendency to give socially desired answers:* respondents are not anonymous → social pressure is high |
| ++ |  |

| **3** Reliability | **Possibility for large samples:** No, usually < 500 (no compensation for noise in the data)  
*Training and Supervision:* Interviewers have to be trained, monitoring is better than with «in home» interviews; monitoring is very good in the teststudio (does interviewer choose respondents correctly, is the interview conducted correctly?)  
*Computer assisted interview:* answers can be checked for plausibility |
| ++ |  |
## Offline Face-to-Face «in hall»

0 = bad; + = medium; ++ = good; +++ = very good

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 4        | **Visual/ audio/ sensory aids and stimuli**  
**+++** |
|          | **Visual aids:** possible       
**Acustic/ audio aids:** possible       
**Tasting:** possible (particularly in teststudio) |
| 5        | **Survey length and complexity**  
**++** |
|          | **Maximal survey length:** 20 minutes (if respondents are previously length can be increased)       
**Complexity:** can be high due to computer assisted programming of questionnaire |
| 6        | **Costs per interview vs. information gained**  
0 |
|          | **Infrastructure needed:** if teststudio       
**Staff needed:** interviewers |
| 7        | **Timing**  
0 |
|          | **Very time intensive** |
### Offline Paper and pencil

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Representativeness</td>
</tr>
<tr>
<td></td>
<td><strong>Coverage</strong> 100% if access to addresses</td>
</tr>
<tr>
<td></td>
<td>Equal probability to be chosen: random selection of household and of person within household → random sample, no geographical clusters</td>
</tr>
<tr>
<td></td>
<td><strong>Response rate</strong> (realised int./attempted int.): usually low</td>
</tr>
<tr>
<td>2</td>
<td>Validity</td>
</tr>
<tr>
<td></td>
<td><strong>Understanding of questions:</strong> possibility for researcher to do a pretest and verify understanding; possibility to provide additional information below the question text or using pop-up windows</td>
</tr>
<tr>
<td></td>
<td><strong>Interviewer bias:</strong> no interviewer</td>
</tr>
<tr>
<td></td>
<td><strong>Tendency to give socially desired answers:</strong> respondents are anonymous → social pressure is low</td>
</tr>
<tr>
<td>3</td>
<td>Reliability</td>
</tr>
<tr>
<td></td>
<td><strong>Possibility for large samples:</strong> &gt; 1000 possible (compensation for noise in the data)</td>
</tr>
<tr>
<td></td>
<td><strong>Training and Supervision:</strong> not necessary, therefore the work of the interviewers as source for mistakes is eliminated</td>
</tr>
<tr>
<td></td>
<td><strong>Computer assisted interview:</strong> no, therefore answers cannot be checked for plausibility, inconsistent answers and mistakes</td>
</tr>
<tr>
<td></td>
<td>There is no guarantee about «who» answers the questionnaire</td>
</tr>
<tr>
<td>Criteria</td>
<td>Evaluation</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
</tr>
</tbody>
</table>
| 4 Visual/ audio/ sensory aids and stimuli | **Visual aids:** possible  
**Acoustic/ audio aids:** not possible  
**Tasting:** not possible |
| 5 Survey length and complexity | **Maximal survey length:** 20 minutes  
**Complexity:** questionnaire must be very simple |
| 6 Costs per interview vs. information gained | **Infrastructure needed:** none  
**Staff needed:** none  
**Costs for** |
| 7 Timing | Rather high |
**Criteria** | **Evaluation**
---|---
1 | **Representativeness**
  + (depends on available addresses and response rate)
  **Coverage** 100% if access to addresses
  **Equal probability to be chosen:** random selection of household and of person within household → random sample, no geographical clusters
  **Response rate** (realised int./attempted int.): usually low

2 | **Validity**
  ++
  **Understanding of questions:** possibility for researcher to do a pretest and verify understanding; possibility to provide additional information below the question text or using pop-up windows
  **Interviewer bias:** no interviewer
  **Tendency to give socially desired answers:** respondents are anonymous → social pressure is low

3 | **Reliability**
  ++
  **Possibility for large samples:** > 1000 possible (compensation for noise in the data)
  **Training and Supervision:** not necessary, therefore the work of the interviewers as source for mistakes is eliminated
  **Computer assisted interview:** no, therefore answers cannot be checked for plausibility, inconsistent answers and mistakes
  There is no guarantee about «who» answers the questionnaire
## Criteria Evaluation

**0** = bad; **+** = medium; **++** = good; **+++** = very good

### Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 4 Visual/ audio/ sensory aids and stimuli                                | **Visual aids**: possible  
**Acustic/ audio aids**: not possible  
**Tasting**: not possible |
| ++                                                                      |                                                                           |
| 5 Survey length and complexity                                           | **Maximal survey length**: 20 minutes  
**Complexity**: questionnaire must be very simple |
| ++                                                                      |                                                                           |
| 6 Costs per interview vs. information gained                            | **Infrastructure needed**: none  
**Staff needed**: none  
**Costs** |
| ++                                                                      |                                                                           |
| 7 Timing                                                                 | **Rather high**                                                          |
| 0                                                                       |                                                                           |
The data collection tool: the questionnaire
Exercise

Please imagine you belong to the following target group:

- You do not work in marketing, market research, food production/processing, retail, wholesale
- You are responsible for food shopping
- You purchase vegetables at least very rarely
- You consume tomatoes at least very rarely
- You are between 18 and 79 years of age
- You live in Switzerland

Fill in the questionnaire and try to answer the following questions:

- Which question types are there with respect to content and format?
- What structure does the questionnaire follow?
- Where are the mistakes?
The data collection tool: the questionnaire

Question types

With respect to content we differentiate between:

• Awareness/ knowledge questions
• Questions about attitudes
• Questions about behaviour

With respect to format we differentiate between:

• Open vs. closed questions (special case: half open questions)
  • Numeric or text
• Direct vs. indirect questions
• Simple questions with one or more than one answers
• Scaled questions
• Maximum difference scaling
The data collection tool: the questionnaire
Question types

Open vs. half-open questions:

Open questions:
- Rich information content
- Not limited to what is already known to the researcher
- Generates in-depth explanations/understanding
- High evaluation effort
- Not necessarily quantifiable

Half-open questions:
- Combine the advantages of open and closed questions: Not limited to existing knowledge and lower evaluation effort
The data collection tool: the questionnaire

Question types

Direct vs. indirect questions:

Direct questions:
- For unproblematic facts and knowledge questions
- Note: Context and social desirability play an important role here

Indirect questions:
- Projective questions:
  - What do you think, how often do Swiss people drink beer?
  - In what follows you see several statements that have been made by other people. Please indicate to what extent you agree with those statements:
- Associative questions: What comes to your mind if ...?
The data collection tool: the questionnaire

Questionnaire structure

• Introduction
• Screening
• Awareness/Knowledge questions
• Attitudes
• Behaviour
• Sociodemographics
The data collection tool: the questionnaire
Pitfalls/mistakes

• Missing screening question on responsibility for food purchase
• Question order: spontaneous association before knowledge question
• Missing context: supermarket in F170A and F300A
• 2 statements in one: F200, item 5
• 2 questions in one: F110
• Question wording: do not suggest any answers!
• Unbalanced scaling/ missing answers: S03, S04
• Missing answers: F140, F150 (discounter)
• Hypothetical questions: be as concrete as possible!
The data collection tool: the questionnaire
10 commandments for question wording

1. Simple, unambiguous wording, that is understood in the same way by everyone
2. No long and complex questions
3. No hypothetical questions
4. No double negation
5. No suggestive questions
6. No questions which presume knowledge that the majority of the respondents does not have
7. Give unambiguous references regarding time and location
8. Answer categories have to be complete (every respondent has to find a suitable answer category).
9. Answer categories cannot overlap
10. Only use unambiguous expressions/expressions which are not clear to everyone have to be explained
Methods for the analysis of quantitative data
Methods for the analysis of quantitative data
Descriptive analysis

F120: When you are in a supermarket and considering to buy some vegetables, which criterion is most important to you and which criterion is least important to you?; Maximum Difference Scaling; for shares < 3% percentage numbers are not displayed; n = 508

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Low importance (1-33)</th>
<th>Middle importance (34-66)</th>
<th>High importance (67-100)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Produced in your region</td>
<td>7%</td>
<td>14%</td>
<td>80%</td>
<td>82</td>
</tr>
<tr>
<td>Produced in your country</td>
<td>6%</td>
<td>16%</td>
<td>78%</td>
<td>80</td>
</tr>
<tr>
<td>Good taste</td>
<td>5%</td>
<td>20%</td>
<td>76%</td>
<td>77</td>
</tr>
<tr>
<td>Impeccable and fresh appearance</td>
<td>25%</td>
<td>20%</td>
<td>54%</td>
<td>62</td>
</tr>
<tr>
<td>Organic or pesticide free</td>
<td>25%</td>
<td>26%</td>
<td>49%</td>
<td>60</td>
</tr>
<tr>
<td>Good price (in rel. to other offers of same veg. cat.)</td>
<td>28%</td>
<td>28%</td>
<td>44%</td>
<td>57</td>
</tr>
<tr>
<td>Traditional, old variety</td>
<td>69%</td>
<td>23%</td>
<td>8%</td>
<td>27</td>
</tr>
<tr>
<td>Special or unfamiliar colour</td>
<td>95%</td>
<td>4%</td>
<td>4%</td>
<td>7</td>
</tr>
<tr>
<td>Special or unfamiliar shape or size</td>
<td>95%</td>
<td>4%</td>
<td>4%</td>
<td>7</td>
</tr>
</tbody>
</table>

www.fibl.org

02 August 2018 120
Methods for the analysis of quantitative data
Descriptive analysis

F130C: At what price per kilogram would you say: “These tomatoes are cheap/ expensive/ too expensive/ too cheap.”?; Price Sensitivity Meter (PSM); filter: from german speaking Switzerland, consumes tomatoes and prefers offer 3 (see picture); [n = 213]
Methods for the analysis of quantitative data

Descriptive analysis

Point of marginal inexpensiveness: Euro 2,25

Point of marginal expensiveness: Euro 4,25

F130C: At what price per kilogram would you say: “These tomatoes are cheap/ expensive/ too expensive/ too cheap.”?; Price Sensitivity Meter (PSM); filter: from german speaking Switzerland, consumes tomatoes and prefers offer 3 (see picture); [n = 213]
Methods for the analysis of quantitative data
Descriptive analysis

F300B: At what price per kilogram would you say: “These Diversifood tomatoes are cheap/ expensive/ too expensive/ too cheap.”?; Price Sensitivity Meter (PSM); filter: from german speaking Switzerland, consumes tomatoes, likes Diversifood and prefers offer 3 (see picture); [n = 186]
Methods for the analysis of quantitative data
Descriptive analysis

Point of marginal inexpensiveness: Euro 2.60
Point of marginal expensiveness: Euro 5.25

F300B: At what price per kilogram would you say: “These Diversifood tomatoes are cheap/ expensive/ too expensive/ too cheap.”?; Price Sensitivity Meter (PSM); filter: from german speaking Switzerland, consumes tomatoes, likes Diversifood and prefers offer 3 (see picture); [n = 186]
Methods for the analysis of quantitative data

Descriptive analysis

Mean = 7.9  Mean = 7.8  Mean = 7.8  Mean = 7.8  Mean = 7.2  Mean = 7.2  Mean = 6.7  Mean = 6.7  Mean = 6.2  Mean = 5.4

F200: Please indicate to what extent you agree with the following statements; single punch per row; for shares < 3% percentage numbers are not displayed; n = 508
Methods for the analysis of quantitative data
Descriptive analysis – word clouds

www.wortwolken.com

F250: Why do you like the idea of 'diversifood'?; single punch question; filter: likes Diversifood; n = 381
Methods for the analysis of quantitative data

Descriptive analysis – Coding of open question

- More varietal diversity, more choice, something for everyone, more fun
- More naturalness, conservation of the natural environment, gentle handling of the natural environment
- More taste, more intense taste
- More taste variety
- More value/income for the farmers, support of farmers
- More independence/ freedom of decision for farmers
- (conservation) of local varieties, local/regional specialities
  - biodiversity
- Independence of large multinational firms
- More healthy
- Less pesticides

F250: Why do you like the idea of 'diversifood'?; single punch question; filter: likes Diversifood; n = 381
Methods for the analysis of quantitative data
Descriptive analysis

Visualization methods

uni-variate
1 variable

continuous
categorial

2 (and more)
variables

2 categorical
variables
Both continuous
1 categorical & 1 continuous
> 2 variables
Methods for the analysis of quantitative data

Regression analysis

Dependent variable:
WTP more for iron enhanced finger millet

Age
Gender
Education
Income
Household size
FM consumption per hhmember
Use of iron supp.
Scepticism

\[ R^2 = xy\% \]
# Methods for the analysis of quantitative data

## Regression analysis

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. error</th>
<th>Lower bound</th>
<th>Upper bound</th>
<th>Test statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>log10(Age)</td>
<td>-1.12</td>
<td>0.13</td>
<td>-1.38</td>
<td>-0.85</td>
<td>-8.33</td>
<td>0.000***</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.08</td>
<td>0.35</td>
<td>0.729</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.009**</td>
</tr>
<tr>
<td>1_12 years</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>13_15 years</td>
<td>-0.10</td>
<td>0.04</td>
<td>-0.17</td>
<td>-0.02</td>
<td>-2.529</td>
<td>0.012*</td>
</tr>
<tr>
<td>&gt; 15 years</td>
<td>-0.20</td>
<td>0.04</td>
<td>-0.28</td>
<td>-0.12</td>
<td>-4.892</td>
<td>0.000***</td>
</tr>
<tr>
<td>log10(Income)</td>
<td>0.29</td>
<td>0.04</td>
<td>0.20</td>
<td>0.37</td>
<td>6.62</td>
<td>0.000***</td>
</tr>
<tr>
<td>log10(Household Size)</td>
<td>-0.74</td>
<td>0.11</td>
<td>-0.96</td>
<td>-0.52</td>
<td>-6.567</td>
<td>0.000***</td>
</tr>
<tr>
<td>log10(per HHmember FM consumption)</td>
<td>-0.05</td>
<td>0.05</td>
<td>-0.14</td>
<td>0.05</td>
<td>-0.99</td>
<td>0.323</td>
</tr>
<tr>
<td>Iron supply myself</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Yes</td>
<td>-0.08</td>
<td>0.03</td>
<td>-0.14</td>
<td>-0.02</td>
<td>-2.46</td>
<td>0.014*</td>
</tr>
<tr>
<td>Scepticism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rather low</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
</tr>
<tr>
<td>Rather high</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.07</td>
<td>0.37</td>
<td>0.709</td>
</tr>
</tbody>
</table>

Adjusted $R^2 = 47\%$
Methods for the analysis of quantitative data
Cluster analysis

Cluster analysis or clustering is the task of assigning a set of objects into groups.

Objects in the same cluster should be more similar to each other than to those in other clusters.

To perform clustering one must define a measure of similarity or distance based on the observed values describing different properties of the objects.
Methods for the analysis of quantitative data
Cluster analysis

- Important distinction between hierarchical and partitional sets of clusters

- Partitional Clustering
  A division data objects into non-overlapping subsets (clusters) such that each data object is in exactly one subset

- Hierarchical clustering
  A set of nested clusters organized as a hierarchical tree
Methods for the analysis of quantitative data
Cluster analysis – partitioning clustering – K-means

Given a number of objects and an initial (randomly chosen) set of cluster centers, assign each object to the closest cluster center.
Methods for the analysis of quantitative data
Cluster analysis – partitioning clustering – K-means

+ Update the coordinates of each cluster center to the average coordinate of the objects associated with it
Methods for the analysis of quantitative data
Cluster analysis – partitioning clustering – K-means

+ Re-assign each gene to the closest cluster centre
Recalculate the co-ordinates of each cluster centre according to the average co-ordinate of the genes associated with it.
Repeat these steps until no reassignment is possible (or maximal number of iterations have been reached).
CHOICE MODELLING
Methods for the analysis of quantitative data
Choice modelling

Model decision processes based on choices
• Stated preference
• Revealed preference

Random utility theory
• The utility from the chosen item is a function of the frequency of choosing that item in repeated choices
Attributes and characteristics

Products are divided into attributes
People are described as a collection of characteristics
Choice experiments to calculate the relative importance of the attributes for different segments (described by characteristics)
The hard work is in the design
Choice experiments

Identify the important attributes (often with qualitative research)
  • Should be all that are of interest

Construct a design of attributes and levels
  • Choice set gets very large, very fast
  • Partial factorial
    • Enable estimation with fewer choices
  • Designs available from computer
    • Usually orthogonal
Analysis

Construct and administer survey

Analyse (by computer)

- often beginning with the multinomial logistic regression model
- Consistent with economic demand theory
- Converts choice frequencies to utility estimates
Example

- Samsung and iphone (branded)
- Attributes are memory (2 levels) and price (2 levels)
- Other attributes could be shape, screen, software availability etc..
- Probably not the size of screws, place of assembly, source of the cobalt etc.
Example

Choice 1) Samsung vs iphone (other attributes identical)
  • Utility Samsung > utility iphone (either Samsung brand has +ve value or
   iphone has −ve value?)

Choice 2) Samsung vs iphone (100 francs cheaper)
  • Utility (iphone+100 Francs) > Utility Samsung
  • Utility Samsung brand < 100 Francs

Choice 3) Iphone (with 16gB extra memory) vs iphone (100 francs cheaper)
  • Utility (iphone +16gB) > utility (iphone+100)
  • Utility (16gB) > 100 Francs > utility Samsung brand
Real life example

Cost levels: 10, 30, 50, and 70 Francs
Habitat levels:

Best

Good

Poor

Worst
Results look like this

<table>
<thead>
<tr>
<th>Zurich</th>
<th>Treatment Group to which Respondents were Allocated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (1)</td>
</tr>
<tr>
<td>Attribute Level</td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>path</td>
</tr>
<tr>
<td></td>
<td>all</td>
</tr>
<tr>
<td>Habitat</td>
<td>worst</td>
</tr>
<tr>
<td></td>
<td>poor</td>
</tr>
<tr>
<td></td>
<td>good</td>
</tr>
<tr>
<td></td>
<td>best</td>
</tr>
<tr>
<td>Cost</td>
<td>10Fr</td>
</tr>
<tr>
<td></td>
<td>30Fr</td>
</tr>
<tr>
<td></td>
<td>50Fr</td>
</tr>
<tr>
<td></td>
<td>70Fr</td>
</tr>
</tbody>
</table>
Results look like this
Contact

Robert Home  robert.home@fibl.org
Hanna Stolz  hanna.stolz@fibl.org
Claudia Meier  claudia.meier@fibl.org

Research Institute of Organic Agriculture FiBL
Ackerstrasse 113 / Postfach 219
5070 Frick
Switzerland
Phone  +41 62 8657-272
Fax  +41 62 8657-273
info.suisse@fibl.org
www.fibl.org