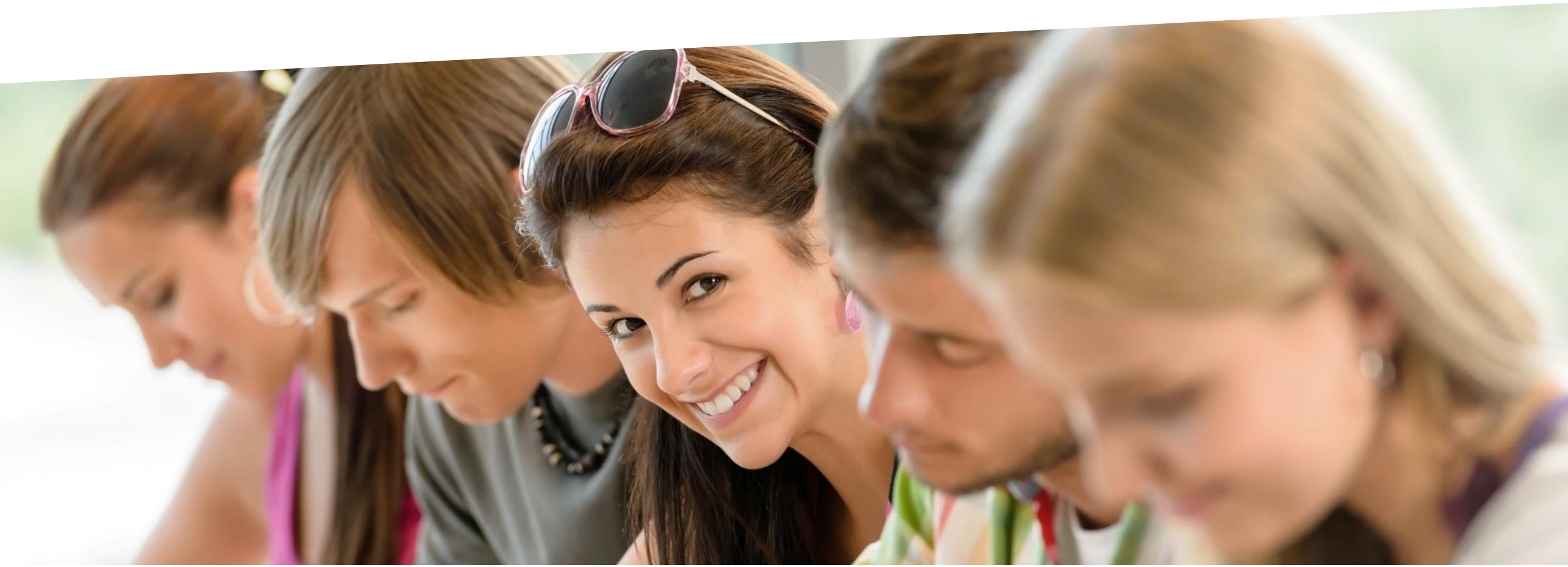
Fachhochschule Graubünden
University of Applied Sciences

# Empirical Social Research

Prof. Dr. Thorsten Merkle / Dr. Ramon Leemann



graubünden Education and Research

# Agenda

- Learning objectives
- Sampling
- Probability sampling
- Non-probability sampling
- Review and self-check questions

FH GR

# Learning objectives

Students are able to …

- ✓ … explain why sampling in business and management research is needed.
- ✓ … distinguish between a range of probability and non-probability sampling techniques.
- ✓ … select and use appropriate sampling techniques for a variety of research scenarios and justify their selection.
- ✓ …assess the representativeness of the sample selected.

FH GR

# Why measure?

- To delineate differences between people, organizations, or any other unit of analysis
- To provide a consistent device for gauging distinctions
- To produce precise estimates of the degree of the relationship between attributes or items we are interested in.

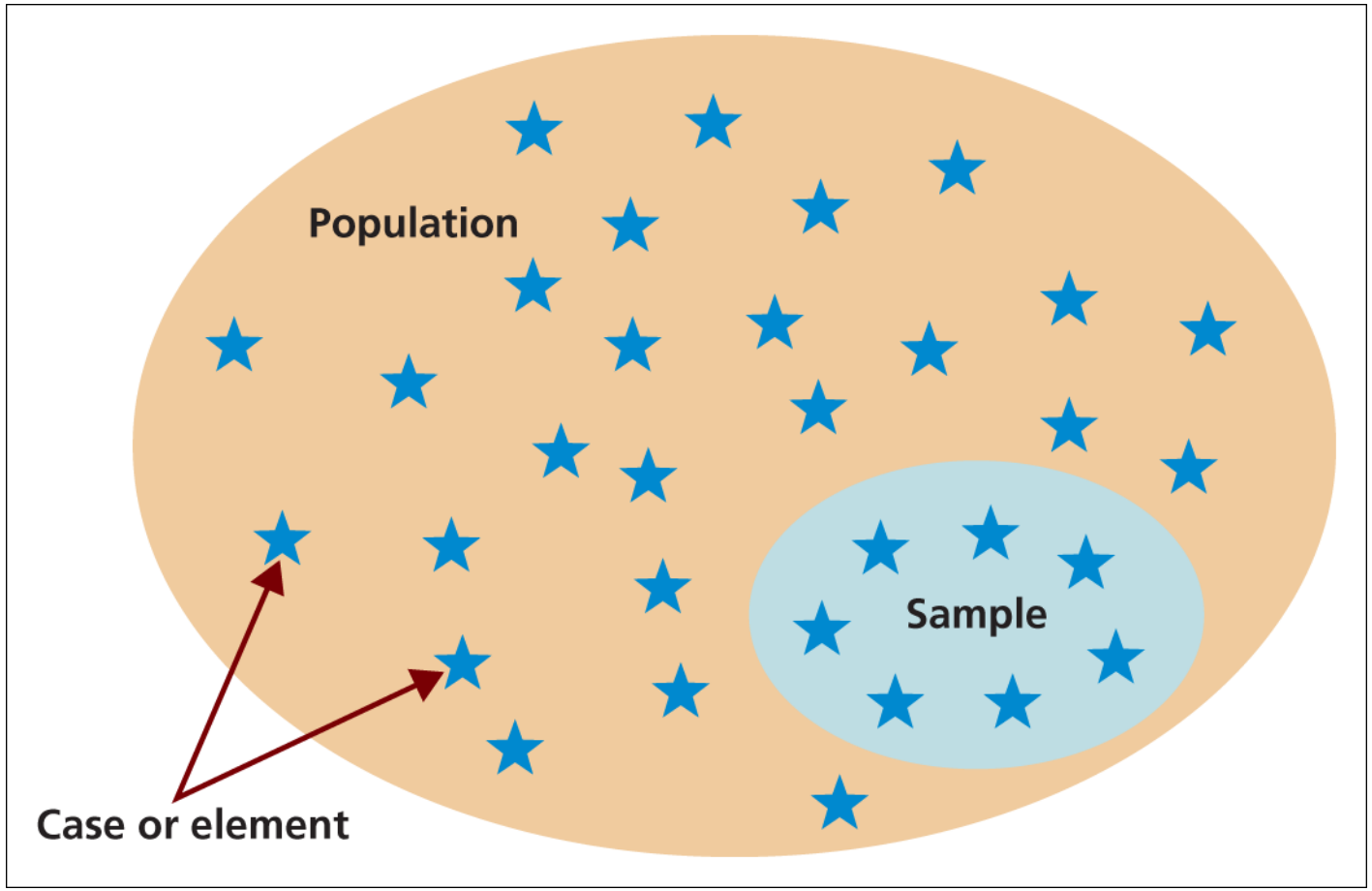
FH GR

# The need to sample

- it would be impracticable for you to survey the entire population;
- your budget constraints prevent you from surveying the entire population;
- your time constraints prevent you from surveying the entire population;
- you have collected all the data but need the results quickly.

FH GR

# Basic terms and concepts




Saunders, et.al. (2012)

# Basic terms and concepts

- Population: the universe of units from which the sample is to be selected
- Sample: the segment of population that is selected for investigation
- Sampling frame: list of all units
- Representative sample: a sample that reflects the population accurately
- Sample bias: distortion in the representativeness of the sample

FH GR

# Basic terms and concepts

- Probability sample: sample selected using random selection
- Non-probability sample: sample selected not using random selection method
- Sampling error: difference between sample and population
- Non-sampling error: findings of research into difference between sample and population
- Non-response: when members of sample are unable or refuse to take part
- Census: data collected from the entire population

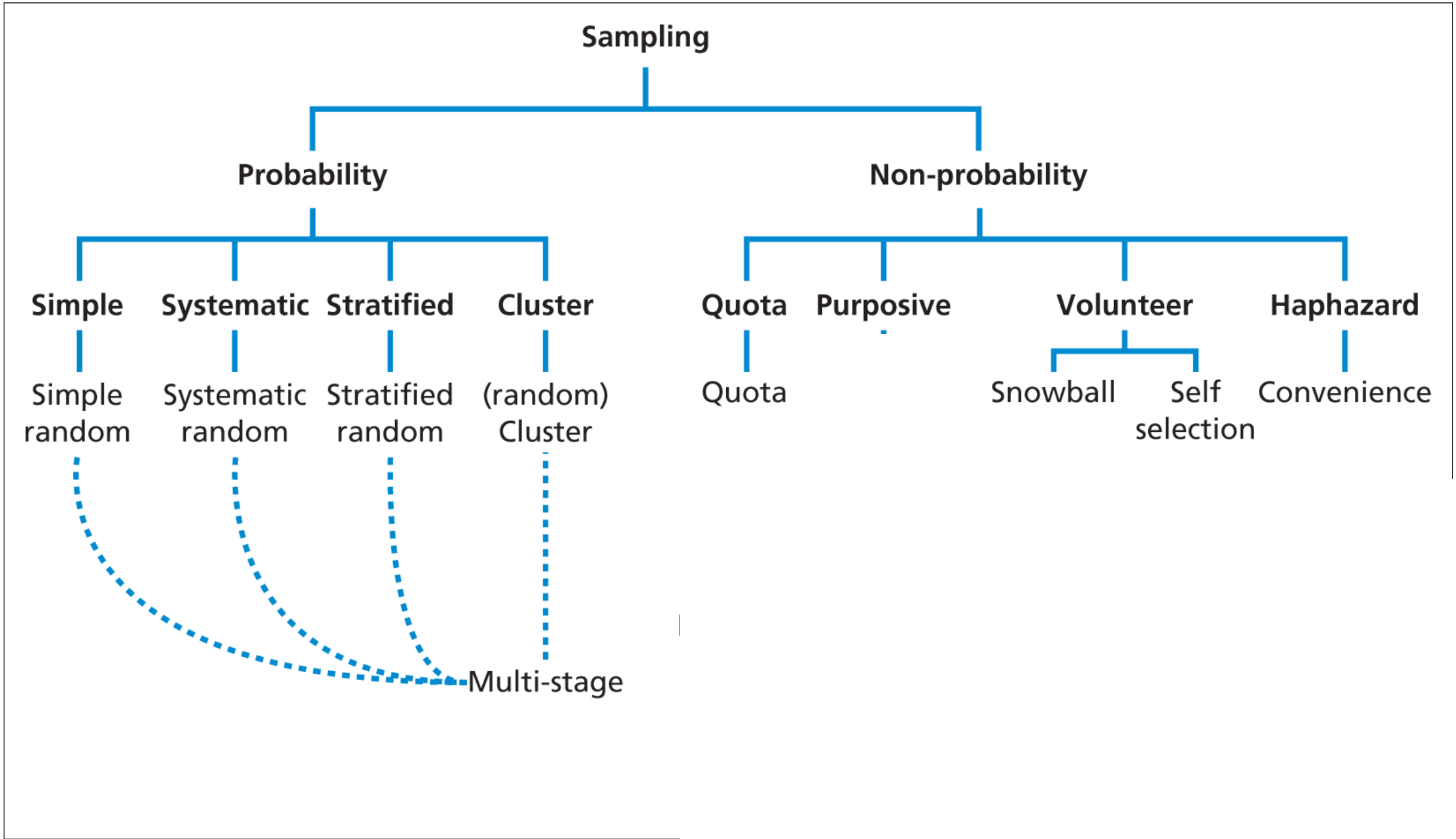
FH GR

# Sampling

# An overview of sampling techniques

The sampling techniques available to you can be divided into two types:

- **probability** or representative sampling;
- **non-probability** or judgemental sampling.


Sampling
Probability
Non-probability
Simple
Systematic
Stratified
Cluster
Quota
Purposive
Volunteer
Haphazard
Simple random
Systematic random
Stratified random
(random) Cluster
Quota
Snowball
Self selection
Convenience
Multi-stage

# Sample size

- Absolute size matters more than relative size
- The larger the sample, the more precise and representative it is *likely* to be ( = **“law of large numbers”)**
- As sample size increases, sampling error decreases
- Important to be honest about the limitations of your sample

FH GR

| | Margin of error | | | |
|---|---|---|---|---|
| *Population* | *5%* | *3%* | *2%* | *1%* |
| 50 | 44 | 48 | 49 | 50 |
| 100 | 79 | 91 | 96 | 99 |
| 150 | 108 | 132 | 141 | 148 |
| 200 | 132 | 168 | 185 | 196 |
| 250 | 151 | 203 | 226 | 244 |
| 300 | 168 | 234 | 267 | 291 |
| 400 | 196 | 291 | 343 | 384 |
| 500 | 217 | 340 | 414 | 475 |
| 750 | 254 | 440 | 571 | 696 |
| 1 000 | 278 | 516 | 706 | 906 |
| 2 000 | 322 | 696 | 1091 | 1655 |
| 5 000 | 357 | 879 | 1622 | 3288 |
| 10 000 | 370 | 964 | 1936 | 4899 |
| 100 000 | 383 | 1056 | 2345 | 8762 |
| 1 000 000 | 384 | 1066 | 2395 | 9513 |
| 10 000 000 | 384 | 1067 | 2400 | 9595 |

# Factors affecting sample size: 1

Time and cost

- after a certain point (n=1000), increasing sample size produces less noticeable gains in precision
- very large samples are decreasingly cost-efficient (Hazelrigg, 2004)

Non-response

- response rate = % of sample who agree to participate (or % who provide usable data)
- respondents and non-respondents may differ on a crucial variable

FH GR

# Factors affecting sample size: 2

Heterogeneity of the population

- the more varied the population is, the larger the sample will have to be

Kind of analysis to be carried out

- some techniques require large sample (e.g. inferential statistics)

# Estimating response rates

- Estimated response rate has an effect on actual sample size required

$$n^a = \frac{n \times 100}{re\%}$$

- $n^a$ is the actual sample size required
- n represents minimum responses required
- re% is the estimated response rate

# Estimating response rates

## Calculation of actual sample size

Jan was a part-time student employed by a large manufacturing company. He had decided to send a questionnaire to the company's customers and calculated that an adjusted minimum sample size of 439 was required. Jan estimated the response rate would be 30 per cent. From this, he could calculate his actual sample size:

$$n^a = \frac{439 \times 100}{30}$$

$$= \frac{43900}{30}$$

$$= 1463$$

Jan's actual sample therefore needed to be 1463 customers. Because of time and financial constraints this was rounded down to 1400 customers. The likelihood of 70 per cent non-response meant that Jan needed to include a means of checking that his sample was representative when he designed his questionnaire.

FH GR

# Sampling error

- Difference between sample and population
- Biased sample does not represent population
  - some groups are over-represented; others are under-represented
- sources of bias
  - non-probability sampling, inadequate sample frame, non-response
- Probability sampling reduces sampling error and allows for inferential statistics

FH GR

# 4 types of probability sampling

- Simple random sample
- Systematic sample
- Stratified random sample
- Multi-stage cluster sample

FH GR

# Simple random sampling

- Each unit has an equal probability of selection
- Sampling fraction: n/N

  where n = sample size and N = population size
- List all units and number them consecutively
- Use random numbers table to select units
- How to generate random numbers…..
  - E.g. use www.randomizer.org
  - Or use Microsoft Excel

FH GR

# Systematic sampling

- Select units directly from sampling frame
- From a random starting point, choose every nth unit (e.g. every 4th name)
- Make sure sampling frame has no inherent ordering – if it has, rearrange it to remove bias

FH GR

# Stratified random sampling

- Starting point is to categorise population into ‘strata’
  (relevant divisions, or departments of companies for example)

- So the sample can be proportionately representative
  of each stratum

- Then, randomly select within each category as for a simple random sample

# The advantages of stratified sampling – example

| Department | Population | Stratified sample | Possible simple random or systematic sample |
|---|---|---|---|
| Sales and marketing | 1,800 | 90 | 85 |
| Finance and accounts | 1,200 | 60 | 70 |
| Human resource management and training | 1,000 | 50 | 60 |
| Technical, research, and new-product development | 1,800 | 90 | 84 |
| Production | 3,200 | 160 | 151 |
| TOTAL | 9,000 | 450 | 450 |

# Cluster sampling

- Useful for widely dispersed populations
- First, divide population into discrete groups (clusters) of units, like geographic areas, or industries, for example. Here, the sampling frame is a complete list of clusters (rather than a complete list of individual cases within the population)
- Second, select a few clusters from the sampling frame (usually using simple random sampling)
- Third, collect data from every case within the selected clusters.

FH GR

# Multi-stage (cluster) sampling

- Development from cluster sampling

- Select clusters (e.g. counties) from the sampling frame (stage 1).
- Divide the selected clusters further into discrete sub-clusters (e.g. electoral wards). The new sampling frame is a complete list of all sub-clusters within the selected clusters.
- Randomly select sub-clusters (stage 2)
- If needed, repeat the division and selection of sub-clusters (stages 3, 4, …)
- Eventually, select cases in the final sub-clusters using a probability sampling method (e.g. systematic sampling).
- Collect data

FH GR

# Qualities of a probability sample

- Representative - allows for generalization from sample to population
- Inferential statistical tests
- Sample means can be used to estimate population means
- Standard error (SE): estimate of discrepancy between sample mean and population mean
- 95% of sample means fall between +/- 1.96 SE from population mean

FH GR

# Limits to generalization

- findings can only be generalized to the population from which the sample was selected
    - be wary of over-generalizing in terms of locality

- time, historical events and cohort effects
    - results may no longer be relevant and so require updating (replication)

# Non-probability sampling

Non-probability sampling is useful …

- … if you don’t have to make statistical inferences.
- … if you can’t or if you don’t need to specify the probability that any case will be included in the sample.

E.g. in in-depth case study research or pilot surveys, that may be followed by probability sampling techniques.

# Non-probability sampling

What is a suitable sample size?

- Sample size is ambiguous, and there are no fixed rules (except for quota samples)
- Sample size depends on the purpose and focus of research.
- In qualitative research, data collection is continued until saturation of data is achieved, i.e. additional data provides few / no new information.

FH GR

# Types of non-probability sampling

Quota sampling

- often used in market research and opinion polls
- relatively cheap, quick and easy to manage
- proportionately representative of a population's social categories (strata)
- but non-random sampling of each stratum's units
- interviewers select people to fit their quota for each category, so the sample may be biased towards those who appear friendly and accessible (e.g. in the street), leading to under-representation of less accessible groups

- Divide population into groups, e.g. age, gender, occupation, etc.
- Decide on a quota for each group
- Select respondents until quota is met

FH GR

# Types of non-probability sampling

Purposive sampling

- Use your judgment to select the cases that will best enable you to answer your research questions.
- Selection with a ‘purpose’, also called judgement sampling
- Cannot be considered representative
- Extreme case sampling
- Maximum variation sampling

# Types of non-probability sampling

Snowball sampling

- Used when it is difficult to identify specific types of population or in order to increase sample size
- researcher makes initial contact with a small group
- these respondents introduce others in their network
- problems of bias: respondents are likely to identify other potential respondents who are similar to themselves, resulting in a homogenous sample

FH GR

# Types of non-probability sampling

Self-selection sampling

- You allow each case, usually individuals, to identify their desire to take part in the research.
- Publicise your need for cases, either by advertising through appropriate media or by asking them to take part.
- Collect data from those who respond.

# Types of non-probability sampling

Convenience / haphazard sampling

- the most easily accessible individuals
- useful when piloting a research instrument
- may be a chance to collect data that is too good to miss

# Error in survey research

Sampling error
- unavoidable difference between sample and population

Sampling-related error
- inadequate sampling frame; non-response
- makes it difficult to generalise findings

Data collection error
- implementation of research instruments
- e.g. poor question wording in surveys

Data processing error
- faulty management of data, e.g. coding errors

FH GR

# Review questions

You want to do research about all students at the University of applied studies in Chur (N = 1’600)

What is the minimum sample size for a 5% margin of error and a 95% confidence level?

What could be an adequate sampling frame?

# Review questions

You want to do research about all inhabitants of Los Angeles (N = 3’900’000)

What is the minimum sample size for a 5% margin of error and a 95% confidence level?

What could be an adequate sampling frame?

FH
GR

# Review questions

You want to do research about all inhabitants of Switzerland  (N = 8’000’000)

What is the minimum sample size for a 5% margin of error and a 95% confidence level?

What could be an adequate sampling frame?

You don’t have access to a sampling frame; which sampling method would you choose?

# Wrap-up

Students …

- ✓ … understand the need for sampling in business and management research
- ✓ … are aware of a range of probability and non-probability sampling techniques
- ✓ … are able to select appropriate sampling techniques for a variety of research scenarios and to be able to justify their selection
- ✓ … are able to use a range of sampling techniques
- ✓ … are able to assess the representativeness of the sample selected

FH GR

# Self-check question 1

You need to undertake a face-to-face interview survey of managing directors of small to medium-sized organisations. From the data you collect you need to be able to generalise about the attitude of such managing directors to recent changes in government policy towards these firms. Your generalisations need to be accurate to within plus or minus 5 per cent. Unfortunately, you have limited resources to pay for interviewers, travelling and other associated costs.

**a** How many managing directors will you need to interview?
**b** You have been given the choice between cluster and multi-stage sampling. Which technique would you choose for this research? You should give reasons for your choice.

# Self-check question 2

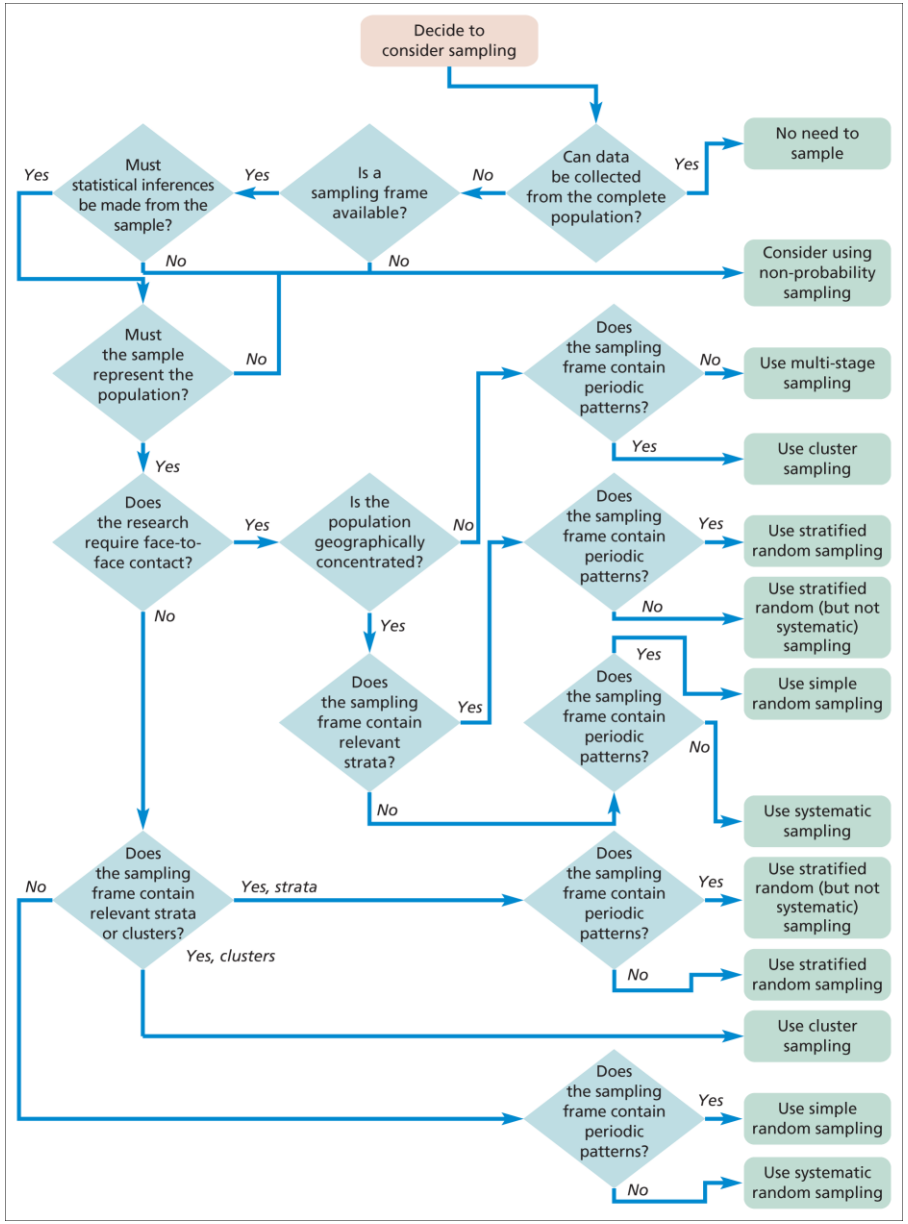
For each of the following research questions it has not been possible for you to obtain a sampling frame. Suggest the most suitable non-probability sampling technique to obtain the necessary data, giving reasons for your choice.

**a** What support do homeless people sleeping on the street believe they require from social services?
**b** Which television advertisements do people remember watching last weekend?
**c** How do employers' opinions vary regarding the impact of new legislation on employee recruitment?
**d** How are manufacturing companies planning to respond to the introduction of road tolls?
**e** Would users of the squash club be prepared to pay a 10 per cent increase in subscriptions to help fund two extra courts (answer needed by tomorrow morning!)?

FH GR

# Roadmap

- Probability-sampling roadmap




**Phase 1**

- Choose sampling frame of relevant discrete groups
- Number each group with a unique number. The first is numbered 0, the second two and so on
- Select a small sample of relevant discrete groups using some form of random sampling

**Phase 2**

- From these relevant discrete groups select a sampling frame of relevant discrete subgroups
- Number each subgroup with a unique number as described in Phase 1
- Select a small sample of relevant discrete subgroups using some form of random sampling

**Phase 3**

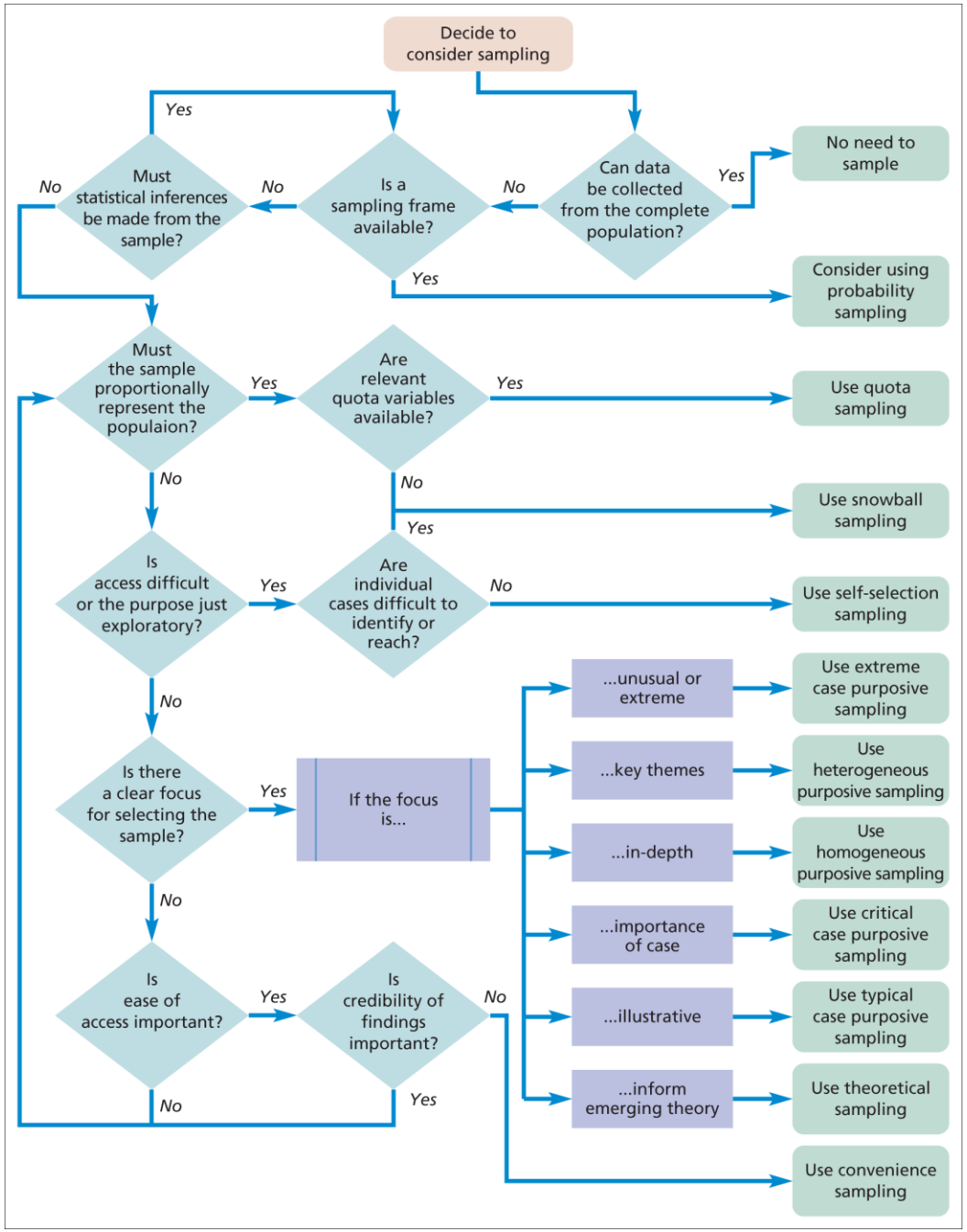
- *Repeat Phase 2 if necessary*

**Phase 4**

- From these relevant discrete subgroups choose a sampling frame of relevant discrete sub-subgroups
- Number each sub-subgroup with a unique number as described in Phase 1
- Select your sample using some form of random sampling

# Roadmap

Non-probability Sampling roadmap



**University of Applied Sciences of the Grisons**
Pulvermühlestrasse 57
7000 Chur
T +41 81 286 24 24
info@fhgr.ch

# Thank you very much for your attention.

Fachhochschule Graubünden
Scola auta spezialisada dal Grischun
Scuola universitaria professionale dei Grigioni
University of Applied Sciences of the Grisons