

● CHAPTER 1 · INTRODUCTORY COURSE

Principles of Survey Research

Methodologically sound. **Practically** applicable.

Market research, survey methodology and their practical application –
from research problem to results report.

Contents

01 Introduction

- 1.1 Market Research and Survey
- 1.2 Types of Market Research

02 Survey: Measurement and Scaling

- 2.1 Introduction
- 2.2 Comparative Scales
- 2.3 Non-Comparative Scales
- 2.4 Latent Constructs
- 2.5 Reliability and Validity

03 Questionnaire

- 3.1 Asking Questions
- 3.2 Overcoming Inability to Answer
- 3.3 Overcoming Unwillingness to Answer
- 3.4 Increasing Willingness of Respondents
- 3.5 Determining the Order of Questions
- 3.6 What's Next?

04 Sampling

- 4.1 Non-probability Sampling
- 4.2 Probability Sampling
- 4.3 Choosing Non-probability vs. Probability Sampling
- 4.4 Sample Size

05 Data Analysis: A Concise Overview of Statistical Techniques

- 5.1 **Descriptive Statistics:** Organizing and Presenting Data
 - 5.1.1 Organizing Qualitative Data
 - 5.1.2 Organizing Quantitative Data
 - 5.1.3 Summarizing Data Numerically
 - 5.1.4 Cross-Tabulations
- 5.2 **Inferential Statistics:** Can the results be generalized to the population?
 - 5.2.1 Hypothesis Testing
 - 5.2.2 Strength of a Relationship in Cross-Tabulation
 - 5.2.3 Relationship between Two (Ratio Scaled) Variables

06 Advanced Techniques of Market Analysis: Some Useful Concepts

- 6.1 Conjoint Analysis
- 6.2 Market Simulations
- 6.3 Segmentation
- 6.4 Perceptual Positioning Maps

07 Reporting Results

1

Introduction

1.1 Market Research and Survey

1.2 Types of Market Research

1

Introduction

1.1 Market Research and Survey

1.2 Types of Market Research

What is Research?

RESEARCH IS

All systematic endeavors and efforts to acquire new knowledge for science or industry.

– Encyclopedia

RESEARCH IS

The search for and collection of information and ideas in response to a specific question.

– Practical definition

Survey

SURVEY

A survey is one of the most popular methods of collecting primary data, in which the researcher interacts with respondents to obtain information about people's attitudes, opinions, knowledge and behaviors.

1

Attitudes & Opinions – how people think and judge

2

Knowledge – what people know and understand

3

Behaviors – what people actually do

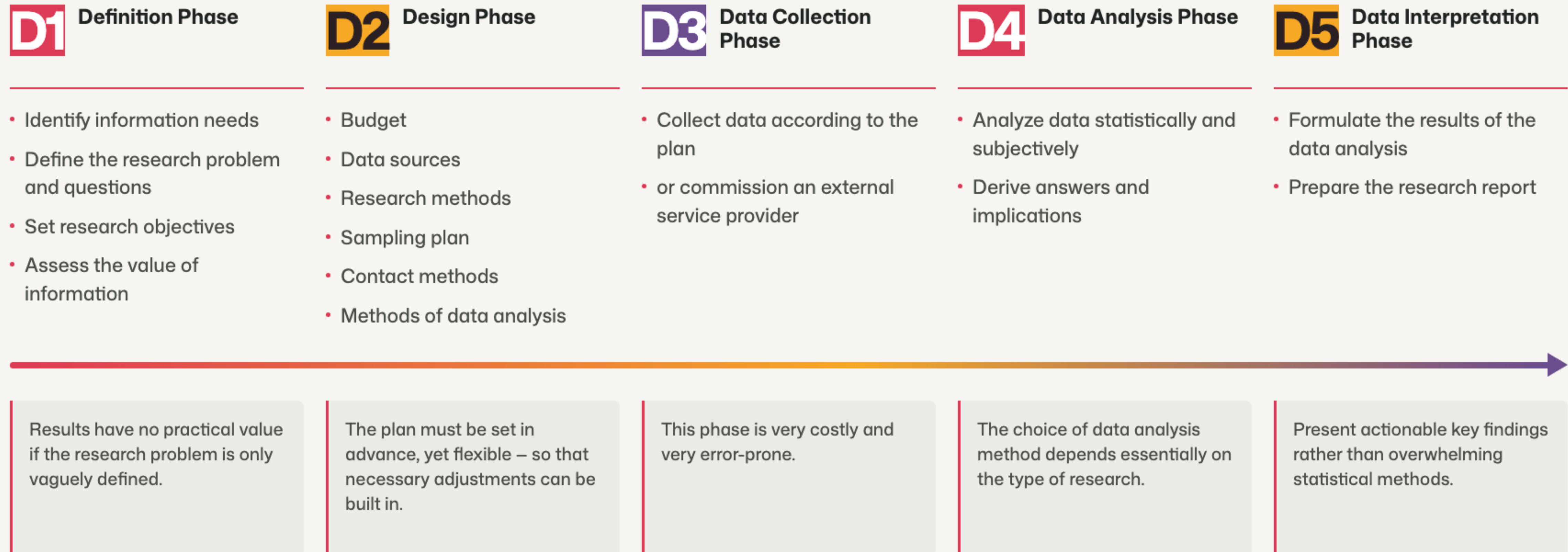
Market Research



Practical Application of Surveys

DISCIPLINE	APPLICATION
Sociology and Political Science	Opinion research, identifying the attitudes of population groups toward socially significant phenomena, events and facts; election research (e.g. the "Sunday question"), ...
Psychology	Personality tests, intelligence tests, identifying individual strengths and weaknesses, psychological stability, cognitive disorders, social influences, ...
Human Resources	Measuring employee satisfaction, loyalty, potential, personality traits, leadership qualities, productivity, professional aptitude, stress resistance, work-life balance, ...
Marketing	Market and consumer research, measuring image perception, preferences, satisfaction and loyalty (NPS), willingness to pay; segmentation, positioning, pricing, advertising tests, website usability, ...
Science (in general)	Studying relationships between two or more variables, factors, phenomena; developing scales and methods for scientific and practical purposes, ...
Education	Knowledge tests (multiple-choice exams), student and teacher evaluation; large-scale educational studies (e.g. PISA), ...
...	... and many further fields of application

Market Research Process – the "5 D's"



When should you **not** start market research projects?

CASE	COMMENT
Vague objectives	When managers cannot agree on what information they need to make a decision. Market research only helps when it investigates a concrete question.
Fixed stance	When the decision has already been made and the study is only meant to "rubber-stamp" a preconceived plan.
Too late	When results are provided too late to still influence the decision.
Poor timing	When a product is in its decline phase, there is little point in researching new product variations.
Insufficient resources	It is not worthwhile to set up a quantitative study as long as no statistically significant sample is feasible – or when the finances are not enough to implement the resulting decisions.
Costs outweigh benefits	The expected value of information should exceed the costs of data collection and analysis.
Results not actionable	When, for example, psychographic characteristics are used that do not help in making concrete decisions.
Information not needed	When decision-relevant information is already available.

1

Introduction

1.1 Market Research and Survey

1.2 **Types of Market Research**

Types of Market Research

CRITERION 1

By Objectives

- **Exploratory**
also diagnostic
- **Descriptive**
- **Causal**
also predictive, experimental

CRITERION 2

By Data Source

- **Primary**
data collected yourself
- **Secondary**
data already available

CRITERION 3

By Methodology

- **Qualitative**
understand
- **Quantitative**
measure

Market Research by Objectives

Uncertain

Uncertainty shapes the type of market research

Certain

LOW CLARITY

Exploratory

also diagnostic

Problem not yet precisely defined; hypotheses are formed.

Analyzing data and actions to understand problems better

- What reasons might lie behind the declining customer satisfaction?
- What keeps first-time buyers from buying again?

Smaller surveys, focus groups, interviews

MEDIUM CLARITY

Descriptive

Phenomena are systematically described and measured.

Collecting and presenting facts: who, what, when, where, how?

- What does the historical sales trend in the industry look like?
- What are consumers' attitudes toward our product?

Larger surveys, observation, etc.

HIGH CLARITY

Causal

also predictive, experimental

Cause-and-effect relationships are tested.

Analyzing cause-and-effect relationships – “What if?”

- Predicting the outcomes of marketing actions
- Effect of advertising spend on sales (how much does one advertising euro yield?)

Experiments, A/B tests, consumer panels

Market Research by Data Source

DATA SOURCE

Primary

- Generating data that **do not yet exist**. These data are analyzed and may, where applicable, be published by the researcher.

TYPICAL METHODS

Surveys, interviews, observation, experiments, ...

DATA SOURCE

Secondary

- Using data collected at **an earlier point in time** for the intended research purpose.

TYPICAL METHODS

Literature research: library, web, database, archive

Market Research by Methodology

METHODOLOGY

Quantitative

- **Involves collecting and evaluating data**
- Requires large amounts of data, uses statistical methods
- Aims for representativeness of the results

TYPICAL METHODS

Larger surveys, observation, etc.

METHODOLOGY

Qualitative

- **Seeks to understand consumer behavior and its causes**
- Focus on individuals and small groups
- Not representative – understanding one perspective, not all

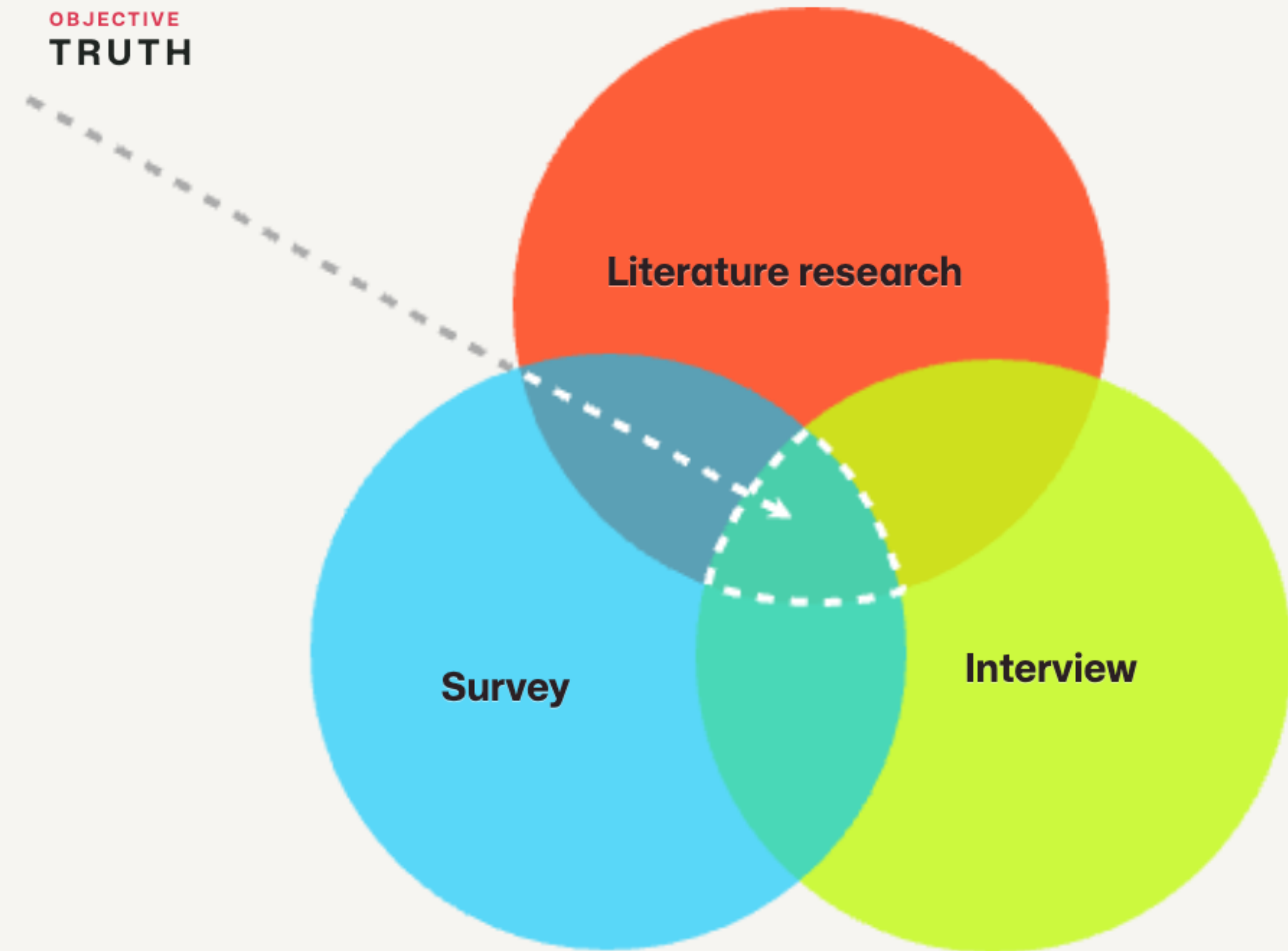
TYPICAL METHODS

Smaller surveys, focus groups, interviews, ...

Triangulation

Triangulation – combining methods within a study on the same topic.

Robson (1998) · Visocky & Visocky (2009)



SO MUCH FOR THE THEORY ...

But in reality everything
is **messier**.



2

Survey: Measurement and Scaling

- 2.1 Introduction
- 2.2 Comparative Scales
- 2.3 Non-Comparative Scales
- 2.4 Latent Constructs
- 2.5 Reliability and Validity

2

Survey: Measurement and Scaling

- 2.1** Introduction
- 2.2 Comparative Scales
- 2.3 Non-Comparative Scales
- 2.4 Latent Constructs
- 2.5 Reliability and Validity

Measurement



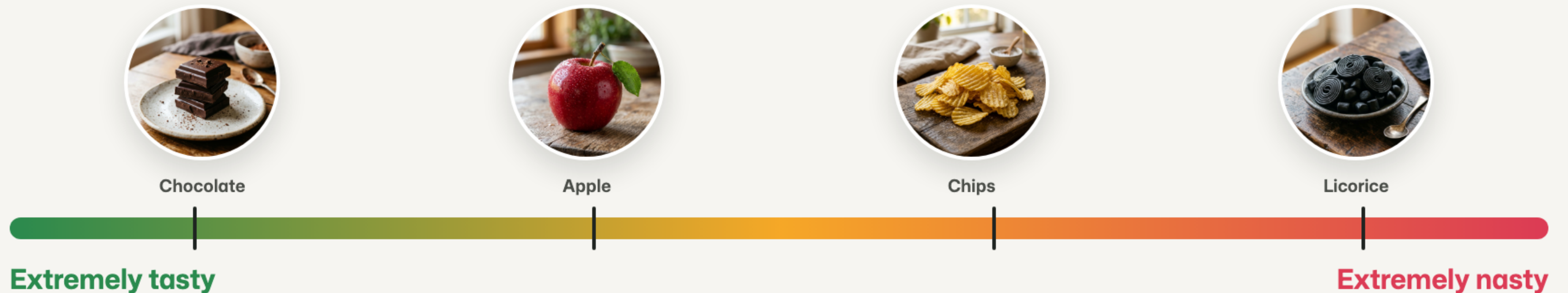
MEASUREMENT

Measurement – assigning numbers or other symbols to characteristics of objects according to a specific, predefined rule.

- 1 One-to-one correspondence of the numbers and the quantities to be measured
- 2 Standardized rules for assigning the numbers
- 3 Rules must **not** vary from object to object or over time

Scaling

Scaling – involves a continuum on which the measured objects are placed.



Primary Scales of Measurement

Nominal

- Numbers merely serve to classify the objects
- non-continuous scale



Ordinal

- Numbers indicate the relative positions of the objects
- but not the magnitude of the difference between them



Interval

quasi-metric

- Differences between objects can be compared
- zero point arbitrary



Ratio

also metric

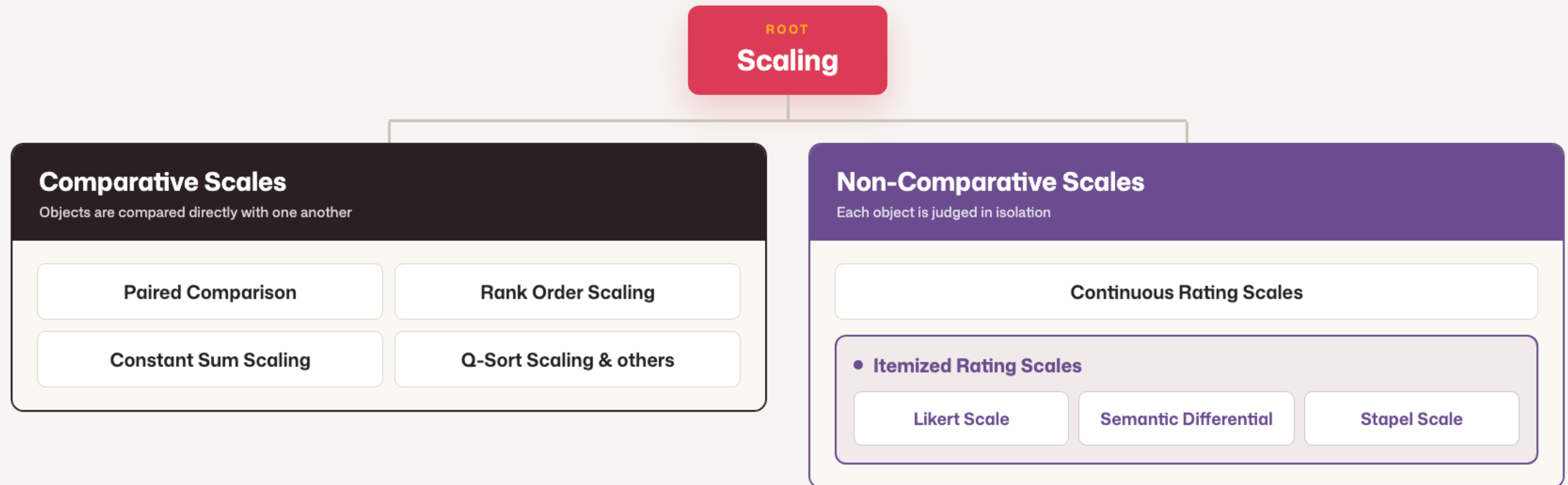
- zero point uniquely fixed
- Ratios of the scale values can be computed



Some Commonly Used Scales in Marketing

SCALE	DESCRIPTION	COMMON EXAMPLES	EXAMPLES FROM MARKETING	STATISTICAL MEASURES	
				DESCRIPTIVE	INFERENTIAL
Nominal Scale	Assignment of numbers to identify and/or classify objects	Passport number, football player's number, gender	Brand, gender, occupation, type of venue	Percentages, mode	Chi-square, binomial test
Ordinal Scale	Numbers describe the rank order of the objects, but not the extent of the differences between them	School grades, position of runners in a marathon	Preference ranking, market position, social class	Percentiles, median	Rank correlation coefficient (Spearman's ρ), Friedman ANOVA
Interval Scale	Allows comparison of the differences between objects; zero point arbitrary	Temperature (Fahrenheit, Celsius)	Attitudes, opinions, purchase intention, customer satisfaction, index numbers	Range, average, standard deviation	Product-moment correlation (Pearson's r), t-tests, ANOVA, regression and factor analysis
Ratio Scale	Zero point is uniquely fixed; allows comparison of both the distances between measured values and their ratios	Length, weight, time, money	Age, revenue, income, costs, market share	Geometric mean, harmonic mean	Coefficient of variation

Classification of Scaling Techniques



Comparison of Scaling Techniques

DIRECT

Comparative Scales

- The measured value of an object results from the **direct comparison** with another object.
- Data can only be interpreted as relative positions – ordinal **ordinal** level of measurement only (rank order).

ISOLATED

Non-Comparative Scales

- Each object is judged in isolation – that is, **independently** of other objects.
- Measurement results are usually treated as **interval** scaled or **metric**.

The choice between the scaling techniques depends on:

Nature of the research question

Variability of the measured value in the population

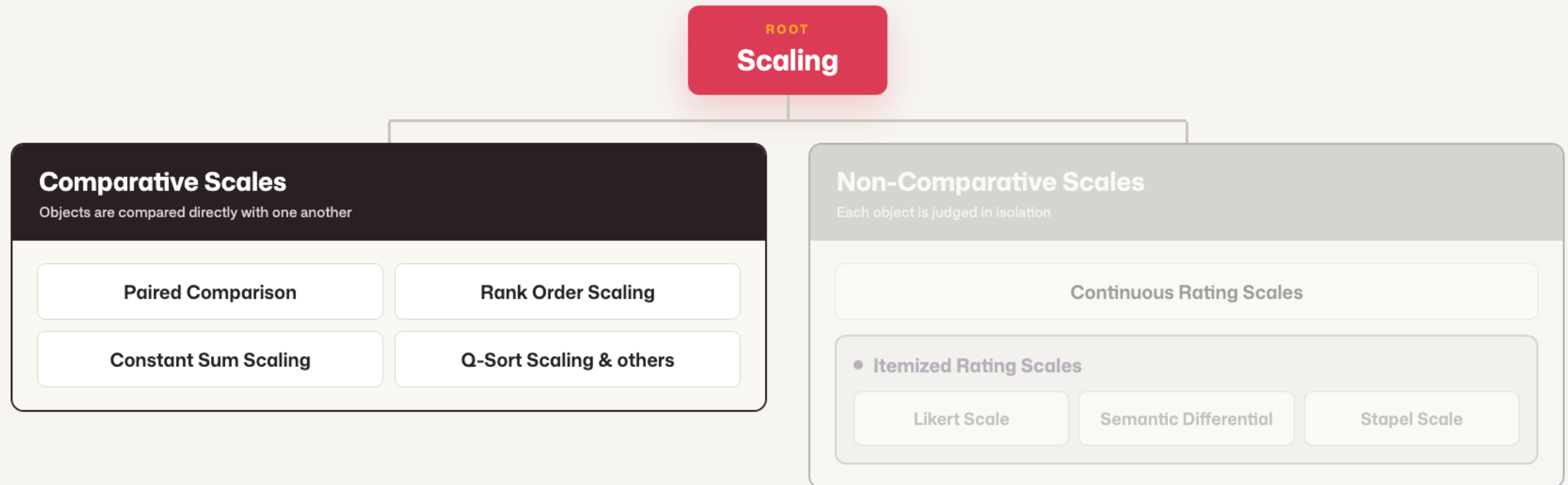
Methods of data analysis

2

Survey: Measurement and Scaling

- 2.1 Introduction
- 2.2 Comparative Scales**
- 2.3 Non-Comparative Scales
- 2.4 Latent Constructs
- 2.5 Reliability and Validity

Classification of Scaling Techniques



Pros-and-Cons of Comparative Scales

+ Pros

- Small differences between objects can be registered
- The same known reference points for all respondents
- Easy to understand and use
- Require fewer theoretical assumptions
- Tend to reduce halo and carryover effects

– Cons

- Only ordinal or rank-order level of measurement → limited choice of statistical methods for data analysis
- Data can only be interpreted as relative positions
- Impossible to generalize beyond the set of objects rated

Comparative Scales: Paired Comparison

Paired Comparison

For each pair of two objects, respondents select the one that in their opinion best fulfills a given criterion.

Below you are presented with ten pairs of beer brands. In each pair, please select the beer you would rather buy.

	WARSTEINER	KÖSTRITZER	OETTINGER	BECKS	PAULANER
Warsteiner		←	←	↑	←
Köstritzer	↑		←	↑	←
Oettinger	↑	↑		↑	↑
Becks	←	←	←		←
Paulaner	↑	↑	←	↑	
# times preferred	3	2	0	4	1

Comparative Scales: Paired Comparison

If these two computers were equal in everything else – which would you prefer?

Pair 3 of 10

COMPUTER ON THE LEFT

1 TB SSD
128 GB RAM

COMPUTER ON THE RIGHT

2 TB SSD
36 GB RAM



Clearly prefer left



Somewhat left



Neutral



Somewhat right



Clearly prefer right

Paired Comparison: Pros-and-Cons

+ Pros

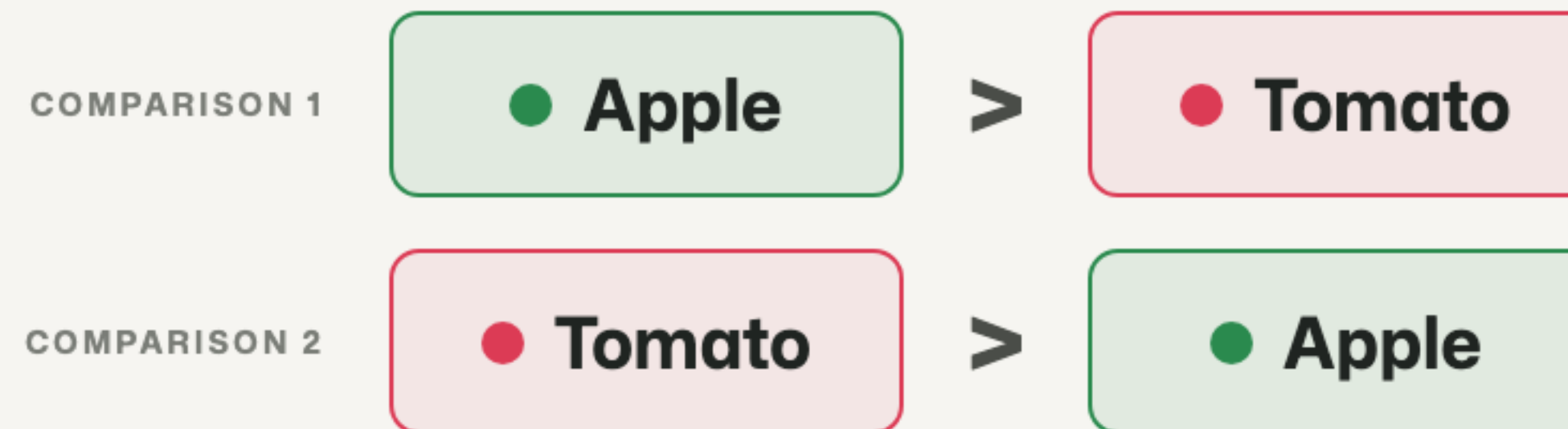
- Direct comparison and unambiguous choice
- Good for blind tests, product comparisons and MDS
- Allows calculating the percentage of respondents who prefer one object
- Rank order can be estimated (assuming transitivity)
- Possible extensions: "no difference" option, graded comparison

– Cons

- Number of comparisons grows faster than the number of objects – for n objects $n(n-1)/2$ comparisons
- Order effects possible (influence of presentation order)
- Preferring A over B does not mean the respondent likes A
- Not very realistic for real choice situations with multiple alternatives
- Violation of the transitivity assumption possible

Violations of transitivity in paired comparison

Same respondent, same pairing – and yet **contradictory** answers:



From "Apple > Tomato" **and** "Tomato > Apple" **no rank order** can be formed – the preferences are **contradictory (intransitive)**.

Violations of transitivity when aggregating preferences

RESPONDENT #1

1 🍏 Apple

2 🍅 Tomato

3 🍊 Orange

RESPONDENT #2

1 🍅 Tomato

2 🍊 Orange

3 🍏 Apple

RESPONDENT #3

1 🍊 Orange

2 🍏 Apple

3 🍅 Tomato

VOTE COUNT

🍏 **Apple** > 🍅 Tomato 2:1

🍅 **Tomato** > 🍊 Orange 2:1

🍊 **Orange** > 🍏 Apple 2:1

Apple > Tomato > Orange > Apple.

Apple is simultaneously the most and the least preferred – the group preferences are inconsistent!

Comparative Scales: Rank Order Scaling

 Rank Order Scaling

Respondents put several objects into an order – based on a particular criterion.

Please arrange the soft-drink brands listed below according to your preferences. To do so, first select the brand you prefer most and assign it rank **1**. Then assign rank **2** to the second-best brand. Continue rating until you have assigned a rank to all brands. The last, least preferred brand must be given rank **5**.

No two brands may receive the same rank.

The preference criterion is entirely up to you. There are no right or wrong answers. Just try to be consistent.

BRAND	RANK
Pepsi-Cola	-----
Coca-Cola	-----
Red Bull	-----
Sprite	-----
7-Up	-----

Rank Order Scales: Example

What would you like for your birthday?

Drag each item into the matching area – rank 1 = most wanted.

TOP-3 WISH LIST

1**Smartwatch****2****Game console****3****Smartphone**

I WOULDN'T WANT IT

**Wireless headphones**

Rank Order Scales: Examples

DRAG & DROP

Select & order top 3

"What matters most to you when shopping? Rank the 3 most important reasons."

AVAILABLE OPTIONS

⋮ Opening hours

⋮ Customer service

⋮ Return policy

YOUR RANKING

1⋮ Selection

2⋮ Online offering

3⋮ Price

RANK MATRIX

Click the rank in each row

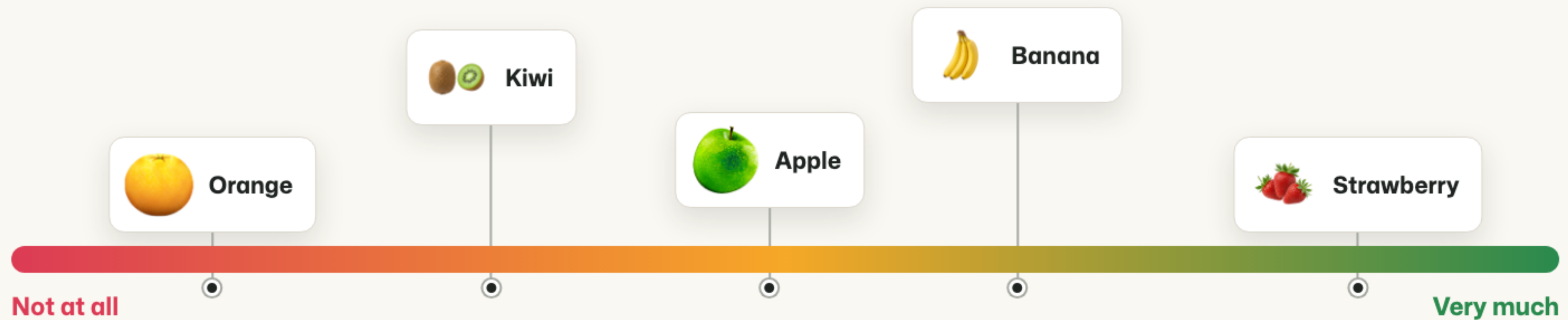
"Put the topic areas into your order (1–6)."

	1	2	3	4	5	6
Communication & Media	☐	☒	☐	☐	☐	☐
Digital media technology	☐	☐	☐	☐	☐	☒
Internet	☐	☐	☒	☐	☐	☐
Office peripherals	☐	☐	☐	☒	☐	☐
IT security	☐	☐	☐	☐	☒	☐
Software & system integration	☒	☐	☐	☐	☐	☐

Rank Order Scales: Example

Which fruit do you like most? – Drag the fruits onto the scale.

Drag & Drop



Rank Order Scales: Pros-and-Cons

+ Pros

- Direct comparison
- More realistic than paired comparisons
 - Number of comparisons is only $(n - 1)$
 - Easier to understand
 - Takes less time
 - No non-transitive answers
 - Data can be converted into paired comparisons
- Good for measuring brand and attribute preferences

– Cons

- Preferring A over B does not mean the respondent likes A
- No zero point – no separation between liking and disliking
- Only ordinal data
- Violation of the transitivity assumption possible (when aggregating)

Comparative Scales: Constant Sum Scaling

 Constant Sum Scaling

Respondents distribute a fixed amount (e.g. points, euros, chips, %) entirely across a set of objects – according to a particular criterion.

Listed below are five attributes of cars. Please distribute **100 points** across these attributes so that the number of points you assign to an attribute reflects its relative importance to you. The more points an attribute receives, the more important it is to you. If an attribute is unimportant to you, assign it **0** points. If one attribute is twice as important as another, assign it **twice as many points**.

ATTRIBUTE	POINTS
Speed	0
Comfort	15
Transmission type (manual/automatic)	5
Fuel (petrol/diesel)	35
Price	45
Sum	100

Constant Sum Scaling: Example of Analysis

ATTRIBUTE	SEGMENT 1	SEGMENT 2	SEGMENT 3
Speed	0	17 ●	53
Comfort	15 ●	23	30
Transmission (manual/automatic)	5	21	10 ●
Fuel (petrol/diesel)	35	12 ●	7
Price	45	27	0
Sum	100	100	100

Constant Sum Scaling: Example

Which features of a rental car are most important to you? Distribute €100 among them.

Sum: **100** / 100

€ ✓



0 €

LEFT

[Reset](#)

Full comprehensive insurance

25 €



Air conditioning

20 €



Navigation (GPS)

15 €



Car max. 3 years old

10 €



Partial comprehensive insurance

10 €



Fuel choice (petrol/diesel)

5 €



CD player

5 €



Radio

5 €



Sunroof

5 €



Area map

0 €

Return instructions

0 €

Constant Sum Scaling: Examples

SLIDER

Drag on scale

"What share of your purchases goes to ...?"

Cosmetics

25 %

Women's fashion

20 %

Men's fashion

15 %

Shoes

25 %

Other

15 %

Total

100 / 100 %

INPUT FIELD

Type in values

"How do you distribute 100% of your monthly budget?"

Housing

40

Food

25

Leisure

20

Other

15

Remaining

0

CLICK BARS

Click on bar

"How is your internet usage distributed?"

private

63 %

work

37 %

Total

100 %

Constant Sum Scaling: Pros-and-Cons

+ Pros

- Can measure small differences between objects without taking up too much time
- Metrically scaled → flexible choice of analysis methods

– Cons

- Results are limited to the list of objects rated – no statements about objects outside the list
- Relatively high cognitive load on respondents, especially with long lists
 - Prone to arithmetic errors (e.g. distribution of 108 or 94 points)

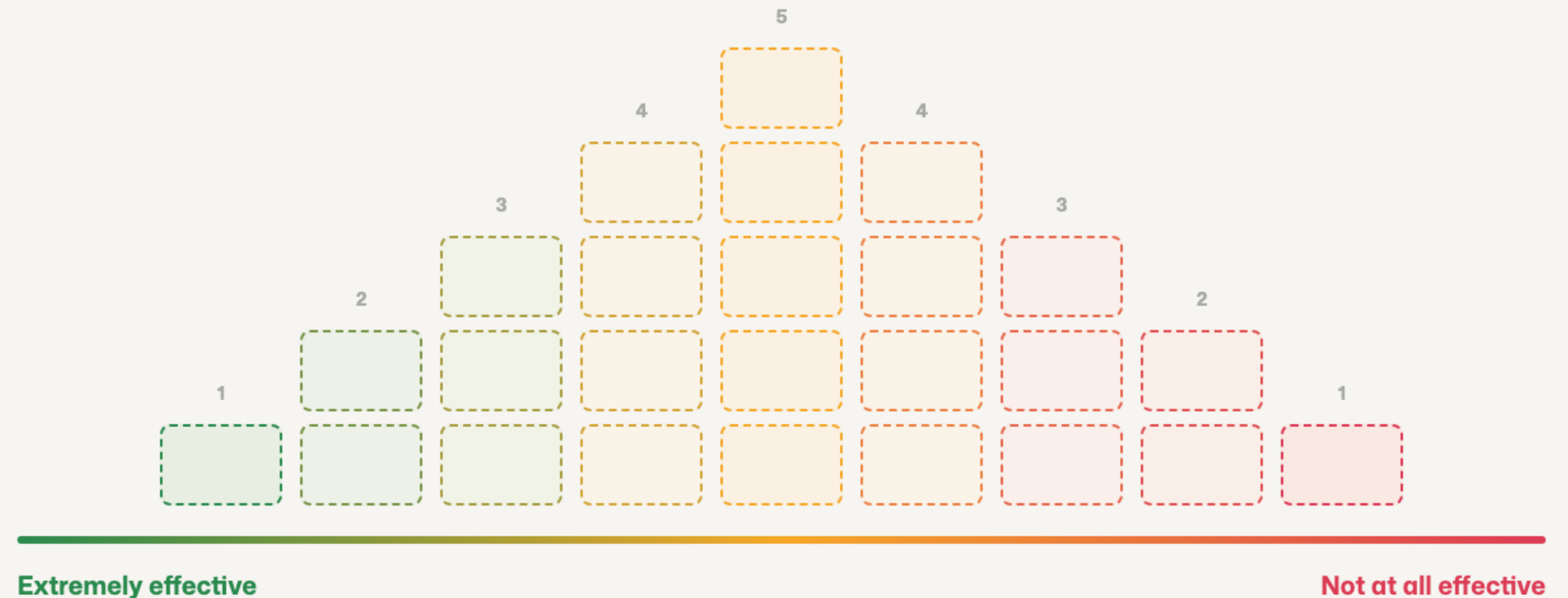
Comparative Scales: Q-Sort Scaling

Q-Sort Scaling

A rank-order procedure in which objects are sorted into piles (with respect to a specific attribute). Used to quickly compare a large number of objects (60–140) against each other.

The number of objects per pile is limited such that all piles together reproduce the shape of a **normal distribution**.

The Ministry of Health has developed **25 measures** for implementation in hospitals. Rank them by their effectiveness against the spread of infection – please only one measure per box.

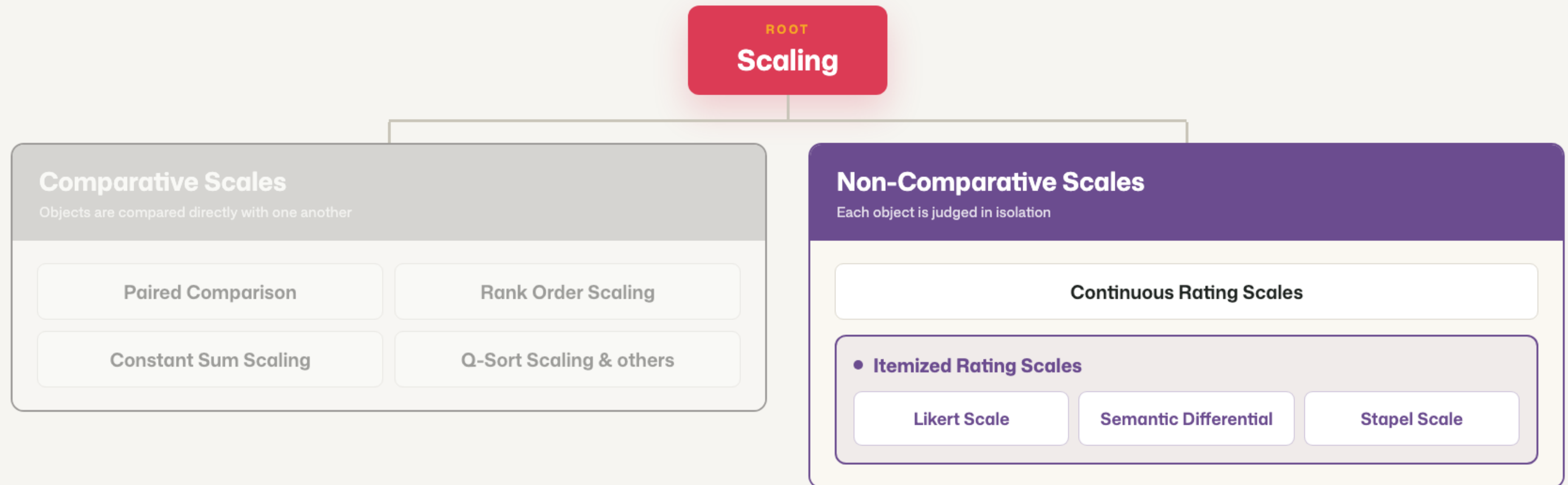


2

Survey: Measurement and Scaling

- 2.1 Introduction
- 2.2 Comparative Scales
- 2.3 Non-Comparative Scales**
- 2.4 Latent Constructs
- 2.5 Reliability and Validity

Classification of Scaling Techniques



Continuous Rating Scale

Continuous Rating Scale

Respondents rate objects by marking a corresponding position on a **line** that runs from one extreme to the other of a given criterion.

How do you rate "Real" as a grocery store?

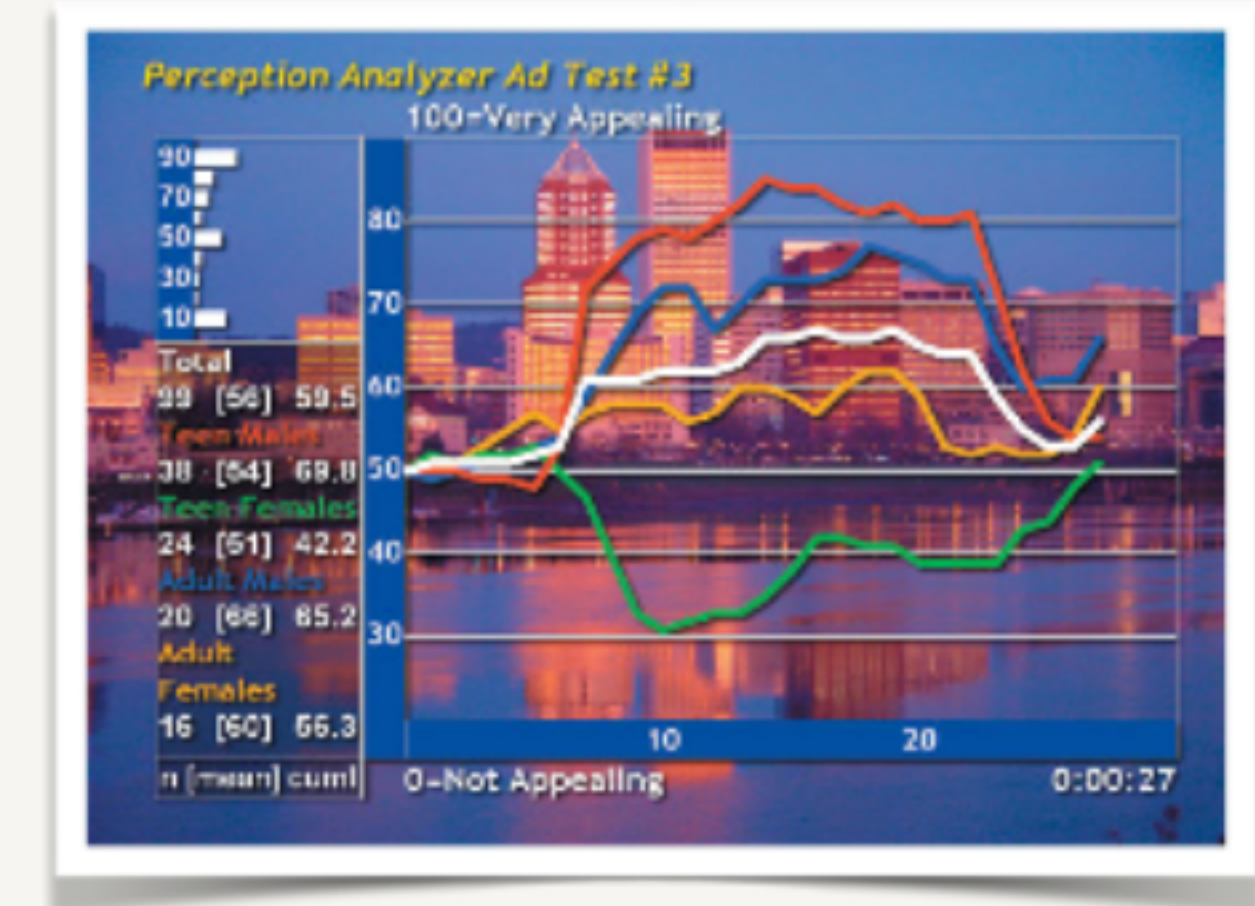


Perception Analyzer

During the presentation of a stimulus – e.g. a TV commercial – each participant turns a dial. This creates a **continuous rating in real time, second by second** – aggregated across all participants and broken down by segments.



Dial handset



Aggregated real-time curves over the spot duration – broken down by segments (0 = "Not Appealing" ... 100 = "Very Appealing").



Participants rate live during the presentation.

Likert Scale

i

Likert Scale

Respondents indicate the extent to which they agree with the listed statements – using a **5- or 7-point scale** that ranges from one extreme to the other.

Below are various statements about "Real". Please indicate how strongly you agree with these statements:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Real sells high-quality goods	☐	☒	☐	☐	☐
Real has poor service REVERSED	☐	☒	☐	☐	☐
I enjoy shopping at Real	☐	☐	☒	☐	☐
Real offers a mix of different brands	☐	☐	☐	☒	☐
The credit policy at Real is terrible REVERSED	☐	☐	☐	☒	☐
I don't like Real's advertising REVERSED	☐	☐	☐	☒	☐
The prices at Real are fair	☐	☒	☐	☐	☐

Important: Statements 2, 5 and 6 are **reversed** in wording. Before data analysis, these scales must be recoded – a higher number should always mean a better attitude.

Likert Scale: Examples

FORMAT · RADIO BUTTONS

Overall, how satisfied are you with our customer service?

☐
Very
dissatisfied

☐
Dissatisfied

☐
Neither

☒
Satisfied

☐
Very
satisfied

FORMAT · SEGMENTED BUTTONS

"The ordering process was easy." – How much do you agree?

Strongly disagree

Somewhat disagree

Neither

Somewhat agree

Strongly agree

FORMAT · NUMBERED SCALE

How important is free return shipping to you?

1

Not at all important

2

3

4

5

Very important

Some Commonly Used Scales in Marketing

CONSTRUCT	1	2	3	4	5
Attitude	Very bad	Bad	Neither good nor bad	Good	Very good
Importance	Not at all important	Unimportant	Neutral	Important	Very important
Satisfaction	Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
Purchase probability (Purchase intention)	Definitely not	Probably not	Undecided	Probably yes	Definitely yes
Purchase frequency	Never	Rarely	Sometimes	Often	Very often
Agreement	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

Scale points from 1 (lowest level) to 5 (highest level).

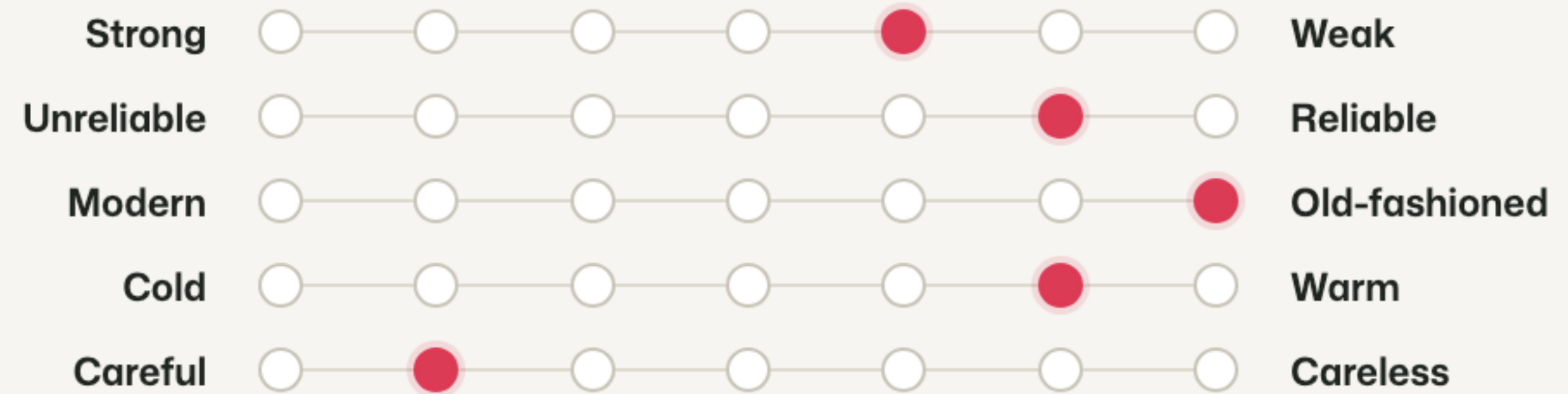
Semantic Differential

Semantic Differential

A bipolar rating scale whose extremes are described by **opposing adjectives**. It allows the measurement of multidimensional attitudes and their **profile representation**.

How do you rate the appearance of “Kaufhof”? Please mark to what extent you lean more toward one pole or the other.

Kaufhof is ...

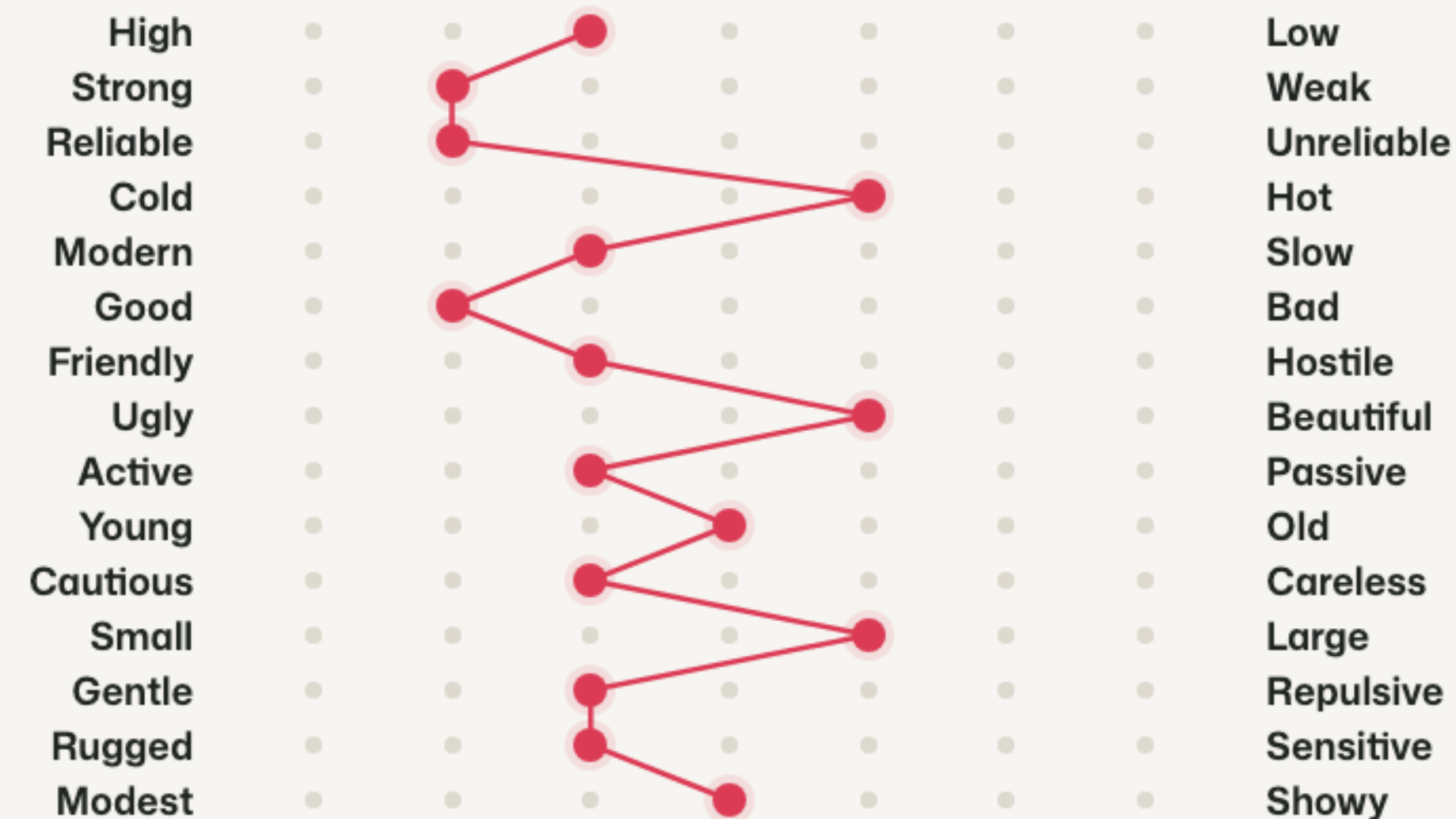


Note: The negative adjectives sometimes appear on the left, sometimes on the right. This makes it possible to check afterwards whether respondents thoughtlessly always marked the same side without reading the adjectives.

Semantic Differential: Profile

Profile representation

Measures self-assessment as well as attitudes toward people or products. **Each point** corresponds to the **mean or median** of the respective scale – connecting the points yields the **profile**.

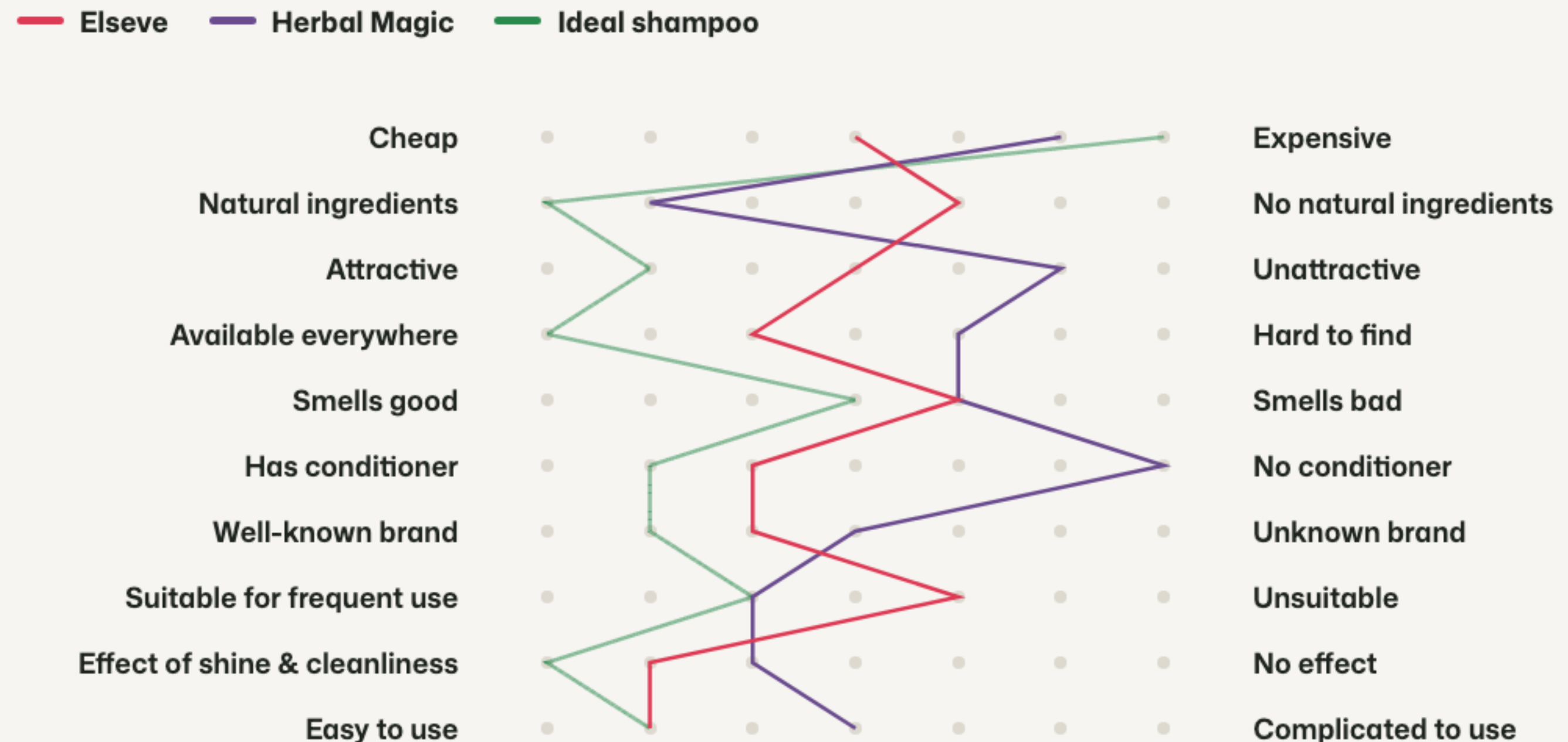


Example profile of an object – seven levels between the opposing poles.

Semantic Differential Scale: Example

Profile comparison

Semantic profiles of the shampoo brands “Herbal Magic” and “Elseve” compared to the **ideal shampoo** from the consumers' point of view.



Stapel Scale

i

Stapel Scale

A **unipolar** rating scale with 10 categories from **−5 to +5**, without a neutral point (0).

It is often used as an alternative to the semantic differential when no meaningful pair of opposing adjectives can be found.

Plus number = the phrase applies, minus number = it does not apply. The larger the magnitude, the stronger.

How accurately do the following phrases describe the store “Real”? For each phrase, choose a number between +5 (fully applies) and −5 (does not apply at all).

+5

+4

+3

+2

+1

High quality

-1

-2

-3

-4

-5

+5

+4

+3

+2

+1

Poor service

-1

-2

-3

-4

-5

Basic Non-Comparative Scales

SCALE	DESCRIPTION	EXAMPLES	PROS	CONS
Continuous Rating Scales	Mark on a continuous line	Reactions to TV commercials	Easy to construct	Manual (non-computer-based) analysis can be very tedious
Likert Scale DISCRETE	Degree of agreement on a scale from 1 (strongly disagree) to 5 (strongly agree)	Measuring attitudes	Easy to understand, use and construct	More time-consuming
Semantic Differential DISCRETE	Bipolar, seven-point rating scale with opposing adjectives at the poles	Brand, product and company image	Versatile	No indication of whether the data are interval-scaled
Stapel Scale DISCRETE	Unipolar ten-point scale from -5 to +5 without a neutral point (0)	Measuring attitudes and image	Easy to construct and to use in telephone surveys	Sometimes confusing and difficult to apply

Constructing Itemized Rating Scales

1

Number of Scale Categories

There is no single optimal number – traditionally, scales with five to nine categories are used.

2

Balanced vs. Unbalanced

In general, the scale should be balanced in order to obtain objective results.

3

Even vs. Odd Number

If a neutral or indifferent response is a viable option for some respondents, an odd number of categories should be chosen.

4

Forced vs. Non-Forced Response

If some respondents may have no opinion, non-forced questions improve the accuracy of the results.

5

Labeling the Scale Points

Not every point needs to be labeled – the key is to avoid ambiguity and to clearly name both poles (and, with an odd number, the midpoint as well).

Number of Scale Categories

1

Number of Scale Categories

There is no single optimal number – traditionally, scales with five to nine categories are used.

More categories capture finer differences – but most respondents can only handle **a few categories**.

Involvement & Knowledge

MORE when respondents are interested in the rating or have deep knowledge of the object.

Nature of the Objects

MORE when fine differences are characteristic of the objects.

Mode of Data Collection

FEWER in telephone interviews, where scales are harder to keep track of.

Data Analysis

FEWER for aggregation & group comparisons · **MORE** for sophisticated, correlation-based statistics.

Balanced or Unbalanced Scales

2

Balanced vs. Unbalanced

In general, the scale should be balanced in order to obtain objective results.

Balanced scale

Very good

Good

Neither good nor bad

Bad

Very bad

2 positive · 1 neutral · 2 negative – symmetrical.

Unbalanced scale

Extremely good

Very good

Good

Adequate

Bad

Very bad

4 positive · 2 negative – skewed toward positive judgments.

Even or Odd Number of Scale Categories

3

Even vs. Odd Number

If a neutral or indifferent response is a viable option for some respondents, an odd number of categories should be chosen.

Odd number – with a midpoint

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

Allows a **neutral** response

Even number – without a midpoint

Strongly disagree

Disagree

Agree

Strongly agree

Forces a **leaning**

The middle option attracts many undecided respondents – and those who are reluctant to reveal their opinion.

This can bias the measures of **central tendency** and **variance**.

Do we want or need “contrast” on controversial attitudes?

Forced or Non-Forced Response?

4

Forced vs. Non-Forced Response

When some respondents cannot have an opinion, non-forced questions improve the accuracy of the results.

Do respondents not want to answer – or do they simply have no opinion?

"Don't know" / "Not applicable"

- ✓ Use for **factual questions** and **knowledge questions**.
- ✗ Not when measuring **attitudes** and **opinions**.

Skip logic

Deliberately steer so that respondents only get questions they can actually **answer** – rather than forcing responses.

Rule of thumb: Questions without a "don't know" tend to yield more accurate data – but only if respondents actually have an opinion.

Labeling the Scale Points

5

Labeling the Scale Points

Not every point needs to be labeled – what matters is avoiding ambiguity and clearly naming both poles (and, with an odd number, the midpoint).

Should every scale point be labeled – or are a few selected points enough?

All or just some?

No clear evidence that labeling **all** points is better than only **selected** ones – the research shows no substantial difference.

Too much confuses

Too many, too finely differentiated labels can confuse respondents when terms can barely be told apart (e.g. "somewhat positive" vs. "fairly positive").

Decisive: clarity

Avoid ambiguity. Clearly name both **poles**, and with an odd number also the **midpoint** – it must be clear which continuum the scale represents.

Little space

Reduced labeling is especially useful with **sliders** or **matrix questions**, where full labeling quickly becomes cluttered.

How Much to Label? – Four Variants

The same question: "How likely are you to buy Product A again?"

All points labeled maximum guidance	☐ Very unlikely	☐ Somewhat unlikely	☐ Neither	☐ Somewhat likely	☐ Very likely
Numbers only minimal labeling	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Only poles labeled endpoints named	☐ Very unlikely	☐	☐	☐	☐ Very likely
Poles + numbers endpoints named, steps numbered	☐ 1 Very unlikely	☐ 2	☐ 3	☐ 4	☐ 5 Very likely

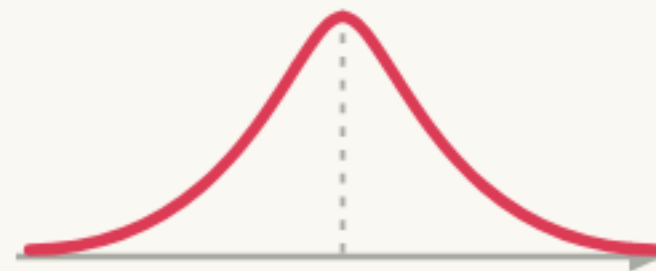
Peaked vs. Flat Response Distribution

How **extreme** the endpoints are worded shapes the form of the response distribution.

EXTREME POLES → PEAKED DISTRIBUTION

extremely satisfied

not at all satisfied



Respondents avoid the extremes – the answers **cluster in the middle**.

MODERATE POLES → FLAT DISTRIBUTION

satisfied

dissatisfied



Respondents also use the endpoints – the answers **spread out more evenly**.

Make clear differences of opinion visible, provoke unambiguous positions

→ **Moderate poles (flatter distribution)**

Set off extreme opinions, make only strong positions visible

→ **Extreme poles (more peaked distribution)**

2

Survey: Measurement and Scaling

- 2.1 Introduction
- 2.2 Comparative Scales
- 2.3 Non-Comparative Scales
- 2.4 Latent Constructs**
- 2.5 Reliability and Validity

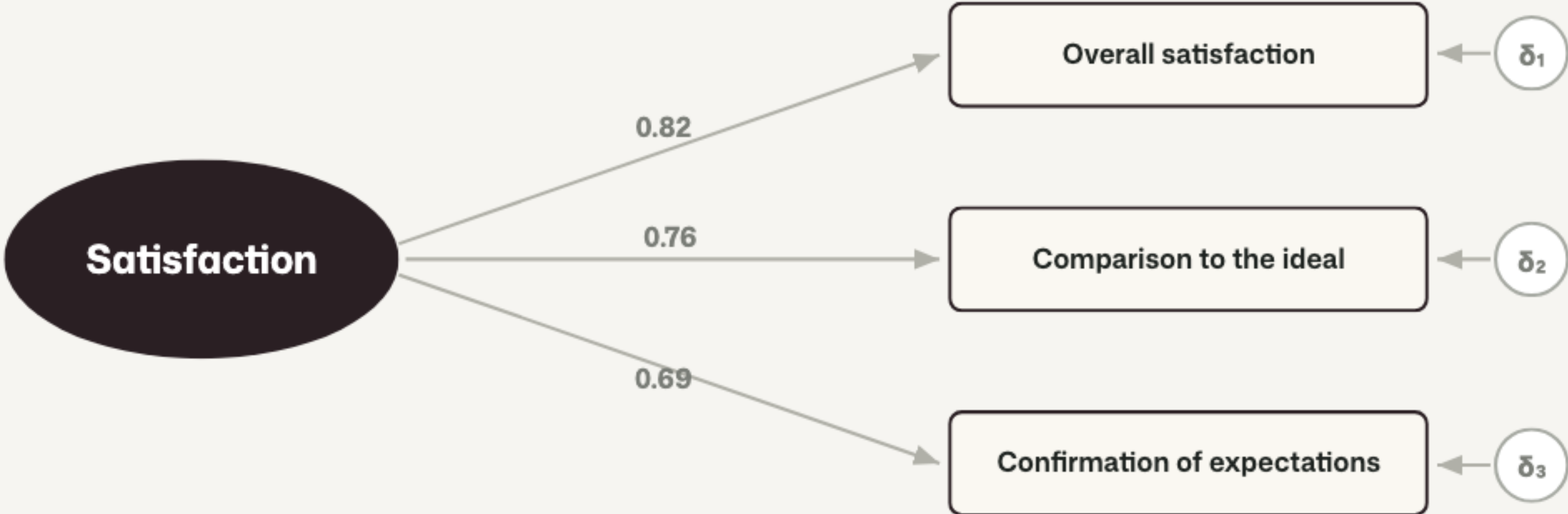
Latent Constructs and Multi-Item Scales

i

Latent construct

A phenomenon (e.g. **customer satisfaction**) that is **not directly** observable or measurable.

That doesn't mean it doesn't "exist" – only that it can be inferred from other, measurable phenomena (**indicators**).

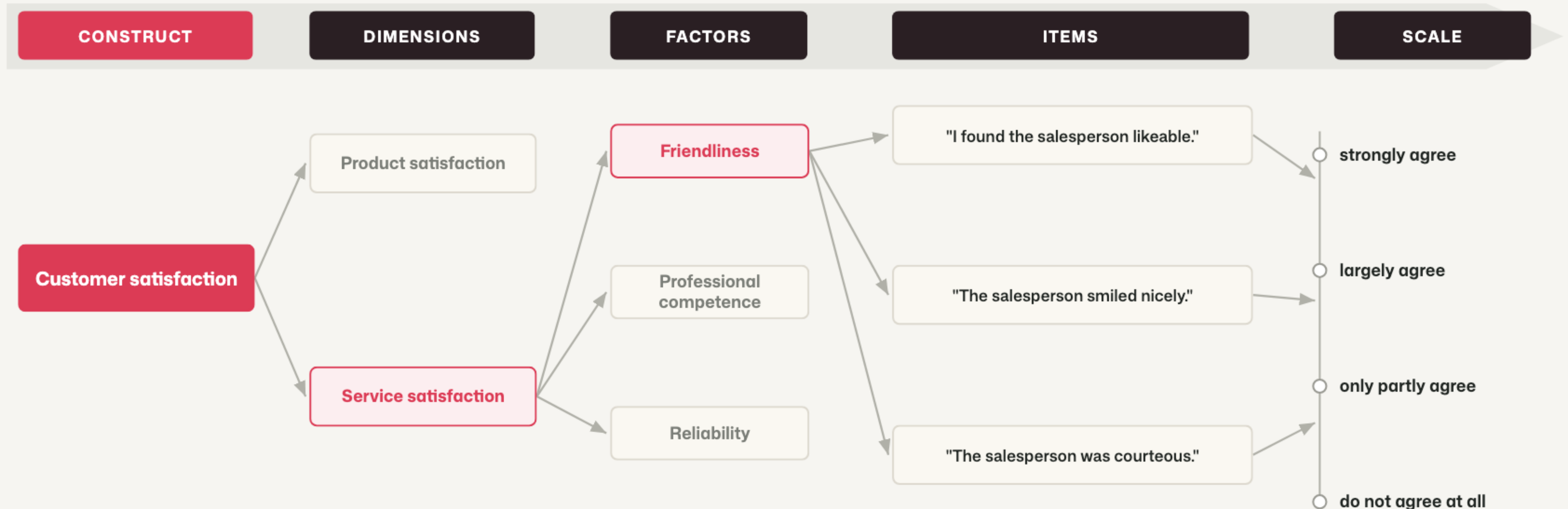


EXAMPLE · HOW "SATISFACTION" IS ASKED

satisfied	☐	☒	☐	☐	☐	☐	☐	dissatisfied
delighted	☐	☒	☐	☐	☐	☐	☐	annoyed
favorable	☐	☐	☒	☐	☐	☐	☐	unfavorable
pleasant	☐	☒	☐	☐	☐	☐	☐	unpleasant
I liked it very much	☐	☐	☒	☐	☐	☐	☐	... not at all
gratified	☐	☒	☐	☐	☐	☐	☐	frustrated
wonderful	☐	☐	☒	☐	☐	☐	☐	terrible

7 items · Cronbach's α = 0.84

Latent Constructs: Hierarchy of Measurement



Multi-Item Scales: Advantages

+ Advantages

- Ability to assess **abstract concepts**
- Different **facets** of the construct can be captured
- **Reduction of data dimensionality** by aggregating many observable phenomena into one model

Examples of latent constructs

Satisfaction

Loyalty

Trust

Service quality

Purchase intention

Brand image

Involvement

Price perception

Usability

Multi-Item Scales: Make or Steal

Make

Develop your own scale



Steal

Adopt ready-made, validated scales

Where do you find ready-made scales?

Bruner (2012): "**Marketing Scales Handbook**", Vol. 6 – a collection of validated multi-item measures.
marketingscales.com/research

AS WELL AS LEADING ACADEMIC JOURNALS

JAMS

JA

JCR

JM

JMR

JR

Secure Customer Index

The **Secure Customer Index** combines three loyalty indicators into **one** metric. Only someone who chooses the top level (5) on **Secure Customer** counts as a Secure Customer on **all three** dimensions – the intersection. Each segment gives the **share (%) of customers**.

Satisfaction in general

5 – very satisfied

4 – somewhat satisfied

3 – neither satisfied nor dissatisfied

2 – somewhat dissatisfied

1 – very dissatisfied

Willingness to recommend

5 – will definitely recommend

4 – will probably recommend

3 – undecided

2 – will probably not recommend

1 – will definitely not recommend

Likelihood of reuse

5 – will definitely reuse

4 – will probably reuse

3 – undecided

2 – will probably not reuse

1 – will definitely not reuse

Secure Customers

Share who are very satisfied / will definitely reuse / will definitely recommend.

Favorable attitude

Share with at least the second-best level on all three dimensions of satisfaction and loyalty.

Vulnerable consumers

Share: somewhat satisfied · undecided · undecided.

At-risk consumers

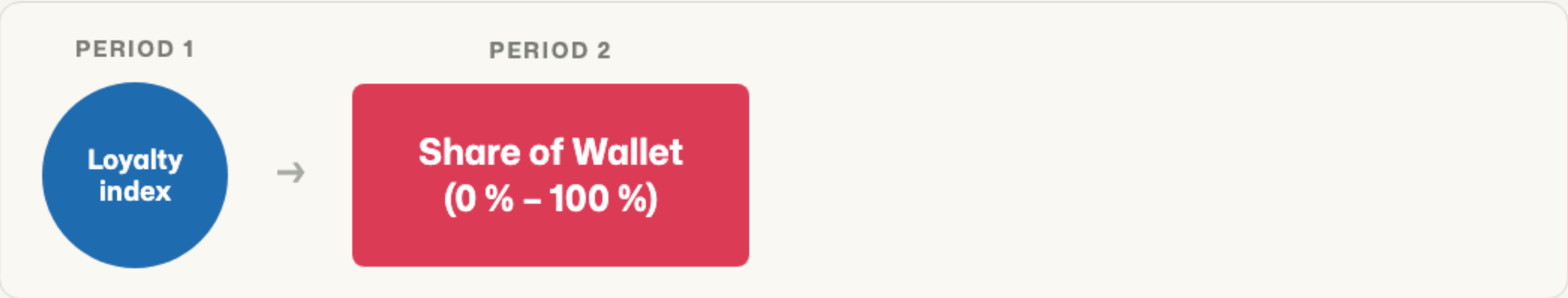
Share: somewhat/not satisfied · will probably or definitely not reuse/recommend.



Source: D. Randall Brandt (1996), "Secure Customer Index", Maritz Research

Extended Secure Customer Index by Burke Inc.

Burke extends the Secure Customer Index by **two additional loyalty dimensions** (five in total) and links the loyalty index measured in **Period 1** with the actual **Share of Wallet** in Period 2 – that is, the share of spending the customer devotes to the brand.



FIVE DIMENSIONS – EACH CAPTURED VIA THE FOLLOWING QUESTION

Overall satisfaction	“How satisfied are you with (BRAND/COMPANY) overall?”
Willingness to recommend	“If you were asked to recommend a company in (INDUSTRY), how likely is it that you would recommend (BRAND/COMPANY)?”
Likelihood to repurchase	“How likely is it that you will continue to use (BRAND/COMPANY)?”
Earned loyalty	“(BRAND/COMPANY) has earned my loyalty.”
Preferred company	“I prefer (BRAND/COMPANY) over all other providers.”



Source: Burke Inc. · <http://www.burke.com/>

2

Survey: Measurement and Scaling

- 2.1 Introduction
- 2.2 Comparative Scales
- 2.3 Non-Comparative Scales
- 2.4 Latent Constructs
- 2.5 Reliability and Validity**

The True-Score Model

The result of a measurement is **not** the true value of a characteristic, but only an **observation** of it.



$$\underset{\text{OBSERVED}}{X_O} = \underset{\text{TRUE}}{X_T} + \underset{\text{SYSTEMATIC}}{X_S} + \underset{\text{RANDOM}}{X_R}$$

- X_O observed value of a characteristic
- X_T the true value of the characteristic
- X_S systematic error
- X_R random error

Reliability and Validity

Reliability

Dependability

- Indicates **how reliably** a measurement instrument measures – i.e. how **consistent** the results are across repeated measurements.
- No random error: $X_R \rightarrow 0 \Rightarrow X_O \rightarrow X_T + X_S$
- The measure is **Cronbach's alpha** ($0 \leq \alpha \leq 1$)
- Values of **$\alpha \geq 0.7$** are considered acceptable

$$X_O = X_T + X_S + X_R$$

Validity

Accuracy

- Indicates **to what extent** a measurement instrument actually measures the matter it is meant to measure.
- That is: to what extent measured differences correspond to **actual** differences between the objects (quality of the measurement).
- No measurement error: $X_S \rightarrow 0, X_R \rightarrow 0 \Rightarrow X_O \rightarrow X_T$

$$X_O = X_T + X_S + X_R$$

Relationship between Reliability and Validity



Reliable
but not valid



Low reliability
low validity



Not reliable
not valid



Reliable and valid
the goal

✓ **Validity implies reliability.**

$$X_O = X_T \Rightarrow X_S = 0, X_R = 0$$

✗ **Non-reliability implies non-validity.**

$$X_R \neq 0 \Rightarrow X_O = X_T + X_R \neq X_T$$

✗ **Reliability does not imply validity.**

$$X_R = 0, X_S \neq 0 \Rightarrow X_O = X_T + X_S \neq X_T$$

Reliability is a **necessary**, but **not sufficient** condition for validity.

”

The purpose of a scale is to enable us to represent respondents with the highest **accuracy** and **reliability**. We cannot have one without the other and still trust our data.

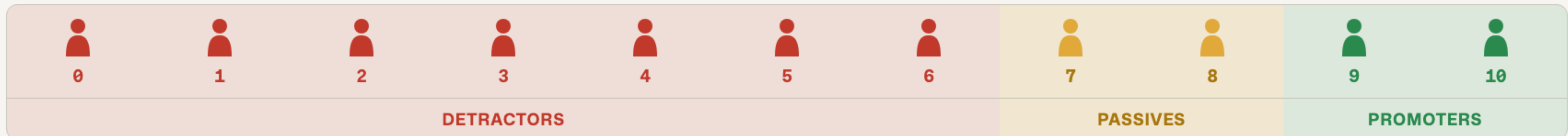
Bart Gamble

Vice President Client Services
Burke, Inc. (2000–2003)



Net Promoter Score® – a predictor of company growth?

„How likely is it that you would recommend company/brand/product X to a friend, relative or colleague?“



$$\text{Net Promoter Score} = \% \text{ Promoters} - \% \text{ Detractors} \rightarrow \text{NPS}$$

-100 % ... +100 %

5-10 %

Average companies

45 %

Companies with open growth potential

50-80 %

Market leaders with high growth potential

Source: Reichheld, Fred (2003) "One Number You Need to Grow", Harvard Business Review

Net Promoter Score®: Warning

”

Although the “**recommendation question**” is by far the best single question for predicting consumer behavior across a range of industries – it is **not** the best question for all industries. That is why companies have to do their homework and **empirically verify**.

the link between the answer and the subsequent consumer behavior for their line of business

Fred Reichheld, 2011



3

Questionnaire

- 3.1 Questionnaire
- 3.2 Asking Questions
- 3.3 Overcoming Inability to Answer
- 3.4 Overcoming Unwillingness to Answer
- 3.5 Increasing Willingness of Respondents
- 3.6 Determining the Order of Questions

Definition & objectives

QUESTIONNAIRE

Questionnaire **A** questionnaire

1

is a formalized list of questions used to gather information from respondents. **“Translate” the information need into a set of unambiguous questions**

2

that respondents are able and willing to answer **Motivate** respondents to take part in the survey and complete it

3

Minimize response error

Questioning Techniques and Questioning Tactics

FORM

Closed vs. Open Questions

- **Closed:** choice from predefined answer options.
 - + easy to analyze, no cognitive stress
 - automatic, unconsidered answers
- **Open:** answer options not predefined.
 - + unlimited possibilities, taxes the memory
 - complex coding, refusal possible

ANALYSIS

closed → easy to evaluate · open → rich, but labor-intensive

APPROACH

Direct vs. Indirect Questions

- **Direct:** the question targets the matter of interest directly.
- **Indirect:** the matter is inferred through an indirect formulation – sparing sensitive or hard-to-articulate topics.

EXAMPLES

direct: "Do you drink alcohol every day?"
indirect: "Which beverages do you prefer with meals?"

Influence of Formulation on the Answer

QUESTION A

"May one smoke while **praying**?"

→ Answer: No

QUESTION B

"May one pray while **smoking**?"

→ Answer: Yes

Same action, different formulation – opposite answers.

Source: Noelle-Neumann & Petersen (1998), p. 192 · n = 2100, p < 0.05

SHARE OF ANSWERS (%)

Yes



No



Unsure



■ "Do you **even** believe in true love?"

■ "Do you believe in true love?"

n = 2100, p < 0.05

What Should Be Considered When Developing a Questionnaire?

1

Is the question necessary?

5

Effort on the respondent's part

9

Easy to complete?

2

Several questions instead of one?

6

Sensitivity of the question

10

Complete & comprehensive?

3

Does the respondent have the information?

7

Explain the aims of the questions

11

Influence of formulation

4

Can the respondent remember?

8

Cultural aspects

3

Questionnaire

3.1 Asking Questions

3.2 Overcoming Inability to Answer

3.3 Overcoming Unwillingness to Answer

3.4 Increasing Willingness of Respondents

3.5 Determining the Order of Questions

3.6 What's Next?

Asking Questions



**"Not every question
deserves an answer."**

Publius Syrus · Rome, 1st c. BC

AVOID ...

Ambiguity & vagueness

Jargon, slang, abbreviations

Double-barreled questions

Leading questions

Implicit assumptions

Implicit alternatives

Treating hypotheses as proof

Generalizations & estimates

Avoid Ambiguity, Confusion and Vagueness

The six W's

Formulate the question in terms of **who, what, when, where, why** and **how**. Especially important: **who, what, when, where**.



EXAMPLE

"Which brand of shampoo do you use?"

ASK INSTEAD

"Which brand or brands of shampoo have you personally used at home during the past month? If you have used more than one, please name all of them."

"Which brand of shampoo do you use?" – what is unclear?

W	ASPECT	WHY UNCLEAR?
Who	Reference person	Unclear whether only the respondent themselves or their entire household is meant.
What	Reference object	Unclear how to answer if several brands are used.
When	Reference period	No reference period given – this morning, this week, or the whole year?
Where	Situation / place	At home, at the gym, on vacation, on a business trip?

Make Answer Options Complete & Unambiguous

For clarity instead of ambiguity: consider all realistic situations and prepare suitable answer options – including "does not apply" and filter routing.

✗ EXAMPLE

"What type of computer do you own?"

- ☐ Windows
- ☐ Mac OS

✓ BETTER

"Which computers do you own?"

- ☐ none
- ☐ Windows
- ☐ Mac OS
- ☐ Other

✗ EXAMPLE

"Are you satisfied with your current car insurance?"

- ☐ Yes
- ☐ No

✓ EVEN BETTER · FILTER ROUTING

1. "Do you have car insurance?" (if no → Question 3)
2. "Are you satisfied with your current car insurance?"

Scales and Answer Options Must Be Unambiguous

Vague frequencies

Words like "rarely," "sometimes" or "often" mean something different to every respondent. Use **concrete, delimited** frequency specifications.

EXAMPLE

"How often do you shop at a supermarket in a typical month?"

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Regularly

ASK INSTEAD

"How often do you shop at a supermarket in a typical month?"

- ☐ less than 1 time
- ☐ 1-2 times
- ☐ 3-4 times
- ☐ more than 4 times

Avoid Jargon, Slang and Abbreviations



Simple words

Use **simple, everyday words** – no technical terms, no jargon. Every respondent must understand the question immediately.



✗ EXAMPLE

"Do you believe that the **distribution** of soft drinks is **adequate**?"

✓ ASK INSTEAD

"Are soft drinks easy to find whenever you want to buy them?"

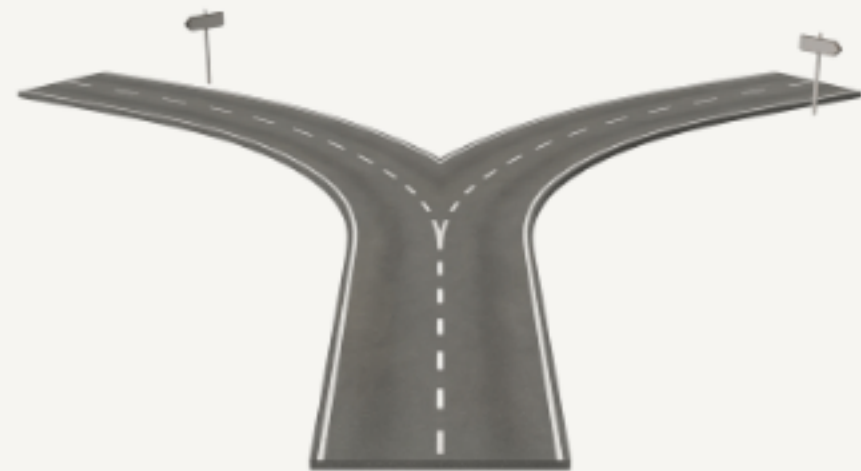
✗ ALSO UNCLEAR

"State your **adjusted net income** for the past year."

Avoid Double-Barreled Questions

One aspect per question

Each question should focus on **just one** aspect. Otherwise you don't know what the answer refers to.



EXAMPLE

"In your opinion, is Coca-Cola **tasty and refreshing**?"

ASK INSTEAD

1. "In your opinion, is Coca-Cola **tasty**?"
2. "In your opinion, is Coca-Cola **refreshing**?"

Avoid Leading

No suggestion

If you already **want** a particular answer, there's no need to ask the question. Leading wording steers the respondent.



EXAMPLE

„**Help the environment** by using cloth shopping bags?"

ASK INSTEAD

"Do you use cloth shopping bags?"

Avoid Implicit Assumptions

Name the consequences

The answer should not depend on **tacit assumptions** about the consequences. Make them explicit.



EXAMPLE

"Do you think the price of milk should be lowered?"

ASK INSTEAD

"Do you think the price of milk should be lowered, **even if this makes the quality of milk worse?**"

Avoid implicit alternatives

Name the alternatives

Implicit alternatives are answer options that were **not explicitly named**. Make the counter-option visible.



EXAMPLE

"Do you like to take the train for short city trips?"

ASK INSTEAD

"Do you like to take the train for short city trips, **or do you prefer to drive?**"

Avoid Treating Beliefs as Real Facts



Facts instead of opinions

Opinions and beliefs often represent the real facts only in a **distorted** way. Ask about the two facts separately.



✕ EXAMPLE

"Do you believe that more highly educated people tend to wear fur clothing more often?"

✓ ASK INSTEAD

1. "What is your level of education?"
2. "Do you wear fur clothing?"

Avoid Generalizations and Estimates

No mental arithmetic

Don't force the respondent to strain their **memory** and their **mathematical skills**.
Ask for the building blocks.



EXAMPLE

"What are the annual per-capita food expenses in your household?"

ASK INSTEAD

1. "How much money does your household spend on food per month?"
2. "How many members does your household have?"

3

Questionnaire

3.1 Asking Questions

3.2 Overcoming Inability to Answer

3.3 Overcoming Unwillingness to Answer

3.4 Increasing Willingness of Respondents

3.5 Determining the Order of Questions

3.6 What's Next?

Overcoming Inability to Answer

1



Is the respondent informed?

Do they even have the necessary information?

2



Can the respondent remember?

Is the information retrievable?

3



Can the respondent articulate?

Can they formulate the answer?

Is the respondent informed?

Respondents often answer questions **even when they are not informed**.

Classic experiment: In response to a question about the fictitious "Consumer Complaints Bureau," **51.9 %** of lawyers and **75 %** of the general public answered – even though no such bureau exists.



✓ REMEDY 1 · FILTER QUESTIONS

Ask in advance about **knowledge** or shopping frequency – e.g. in a study of 10 shops.

✓ REMEDY 2 · "DON'T KNOW"

Offer a "**Don't know**" answer option instead of forcing answers.

Can the respondent remember?

Recall errors

Poor recall leads to **omission**, **telescoping** and **creation**. Ask about **typical behavior**, not exact counts.



Aided recall: "Which of the following brands were advertised yesterday?" (with a list) instead of "Which commercials do you remember?"

EXAMPLE

"How many bottles of soft drinks have you consumed in the past four weeks?"

ASK INSTEAD

"How often do you drink soft drinks in an average week?"

- ☐ less than 1×
- ☐ 1–3× per week
- ☐ 4–6× per week
- ☐ 7× or more

Can the respondent articulate it?

Offer aids

Anyone who **cannot formulate** their answer skips the question or drops out. Offer **pictures, diagrams or descriptions** to choose from.



PROBLEM

When asked to describe the **atmosphere** of a department store, many respondents struggle to formulate their answer.

SOLUTION

When respondents are presented with **alternative descriptions** of the atmosphere, they can select the one they like most.

3

Questionnaire

- 3.1 Asking Questions
- 3.2 Overcoming Inability to Answer
- 3.3 Overcoming Unwillingness to Answer**
- 3.4 Increasing Willingness of Respondents
- 3.5 Determining the Order of Questions
- 3.6 What's Next?

Overcoming Unwillingness to Answer

MOST RESPONDENTS DISLIKE ...

investing a lot of **time and effort** in answering

answering questions that seem **inappropriate** in the context

disclosing information that seems **not useful** to them

revealing sensitive information

THREE REMEDIES

1 Clarify the context

2 Explain the purpose

3 Reduce the effort

Reduce the effort

Reduce the effort

Minimize the effort required to answer.
Instead of recalling freely, let respondents
select from a **ready-made list**.

EXAMPLE

“Please name all the departments where you shopped during your last visit to the department store.”

ASK INSTEAD

“Please tick all the departments where you shopped most recently:”

- ☐ Women’s clothing
- ☐ Men’s clothing
- ☐ Children’s clothing
- ☐ Cosmetics ...
- ☐ Other (please specify) _____

Clarify the context



Provide context

Some questions seem **inappropriate** in the wrong context. Introduce them with an explanatory statement.

Questions about **hygiene habits** seem normal in a medical survey – but out of place in a survey about fast-food restaurants.



INTRODUCTORY STATEMENT

“As a fast-food restaurant, we strive to provide our customers with a clean and hygienic environment. That’s why we’d now like to ask you a few questions about your hygiene habits.”

Explain the purpose



Legitimize the purpose

Explain **why** the information is needed – otherwise it seems intrusive.

Why would a cereal maker be interested in respondents' **age, income and occupation**?



LEGITIMIZE THE INFORMATION REQUEST

“To understand how the consumption of breakfast cereals differs between people of various ages, incomes and occupations, we need the following information from you ...”

3

Questionnaire

- 3.1 Asking Questions
- 3.2 Overcoming Inability to Answer
- 3.3 Overcoming Unwillingness to Answer
- 3.4 Increasing Willingness of Respondents**
- 3.5 Determining the Order of Questions
- 3.6 What's Next?

Handling sensitive topics

- 1 Place sensitive topics at the **end** of the questionnaire
- 2 Introduce with a statement that the behavior is only of interest **in general**
- 3 Phrase questions in the **third person** (as if they concern other people)
- 4 Hide the question within a **group of other questions**
- 5 **Provide** answer options

INSTEAD OF ASKING FOR SPECIFIC DETAILS
OR FIGURES

Sensitive topics

💰 Money

Private & family life

Political & religious views

...

3

Chapter

3.1 Questionnaire

3.2 Asking Questions

3.3 Overcoming Inability to Answer

3.4 Overcoming Unwillingness to Answer

3.5 Increasing Willingness of Respondents

3.6 Determining the Order of Questions

Three principles

1

Determining the Order of Questions

Opening questions **Should be** interesting, easy and non-threatening

2

– they determine whether respondents continue.

Type of information **Rule of thumb: ask the** research-relevant **information first, then the** classification **and lastly the** identification information

3

•

Difficult questions **Place sensitive, embarrassing, complicated or tedious questions as far back**



Funneling and Skip Logic



Funneling

General before Specific

1 GENERAL FIRST

"Which aspects play an important role for you when choosing a department store?"

2 THEN SPECIFIC

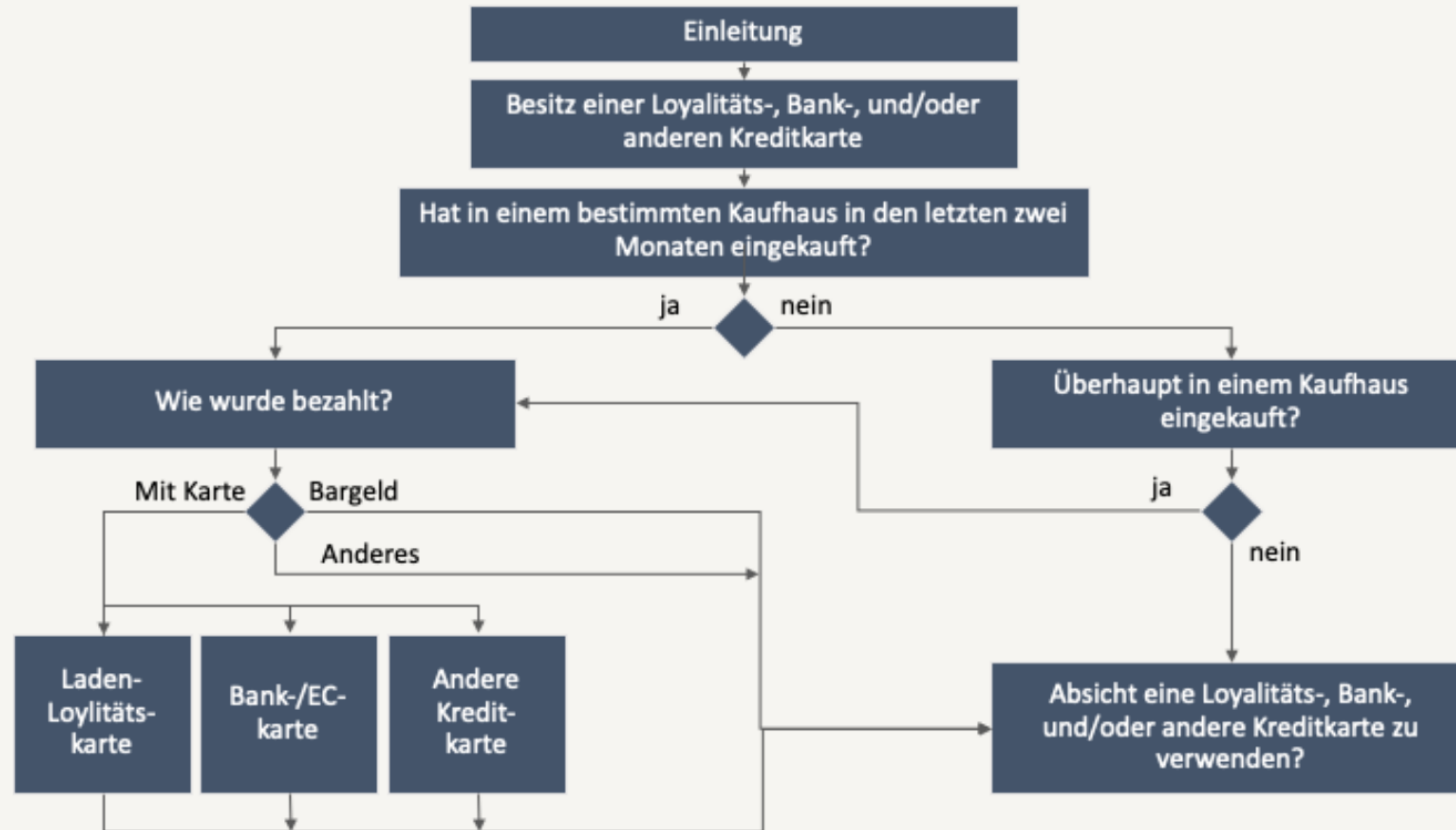
"How important is the convenience of the location to you when choosing a department store?"

Skip Logic

Logical Arrangement

- Place the question branched to as **close** to the triggering question as possible.
- Arrange branching so that respondents **cannot anticipate** which additional information will be asked.

Example: Flowchart of a Questionnaire



3

Questionnaire

- 3.1 Asking Questions
- 3.2 Overcoming Inability to Answer
- 3.3 Overcoming Unwillingness to Answer
- 3.4 Increasing Willingness of Respondents
- 3.5 Determining the Order of Questions
- 3.6 What's Next?**

Design a Convincing Introduction

1

Spark respondents' interest

2

Explain the reasons and goals

3

Ask respondents for help

4

Emphasize that their support is valuable

5

State how long the survey takes

6

Emphasize anonymity

7

Provide incentivesPreferably **non-monetary** incentives.

Pretest! Pretest! Pretest!

Test the questionnaire on a small sample before deployment – and check **every** aspect:

Content of the questions

Wording / phrasing

Order

Form & layout

Difficulty of the question

Instructions

Analysis methods

Recap

FLOWCHART

Develop a **flowchart** of the required information – starting from the (market) research problem.

- Once the sequence is laid out, the **connections** become clear.
- Align the collected data with the **information needs**.
- Set a **clear objective** for each area – the questions follow from it.

PUT ON THE "CRITIC'S HAT"

Go back to the flowchart and ask about each piece of information:

"Do I really need to know this – and do I know what I'll do with it?"

... rather than "Nice to know, but I don't really need it."

4

Sampling

- 4.1 Non-probability Sampling
- 4.2 Probability Sampling
- 4.3 Choosing Non-probability vs. Probability Sampling
- 4.4 Sample Size

THE WORLD'S MOST FAMOUS HEADLINE ERROR

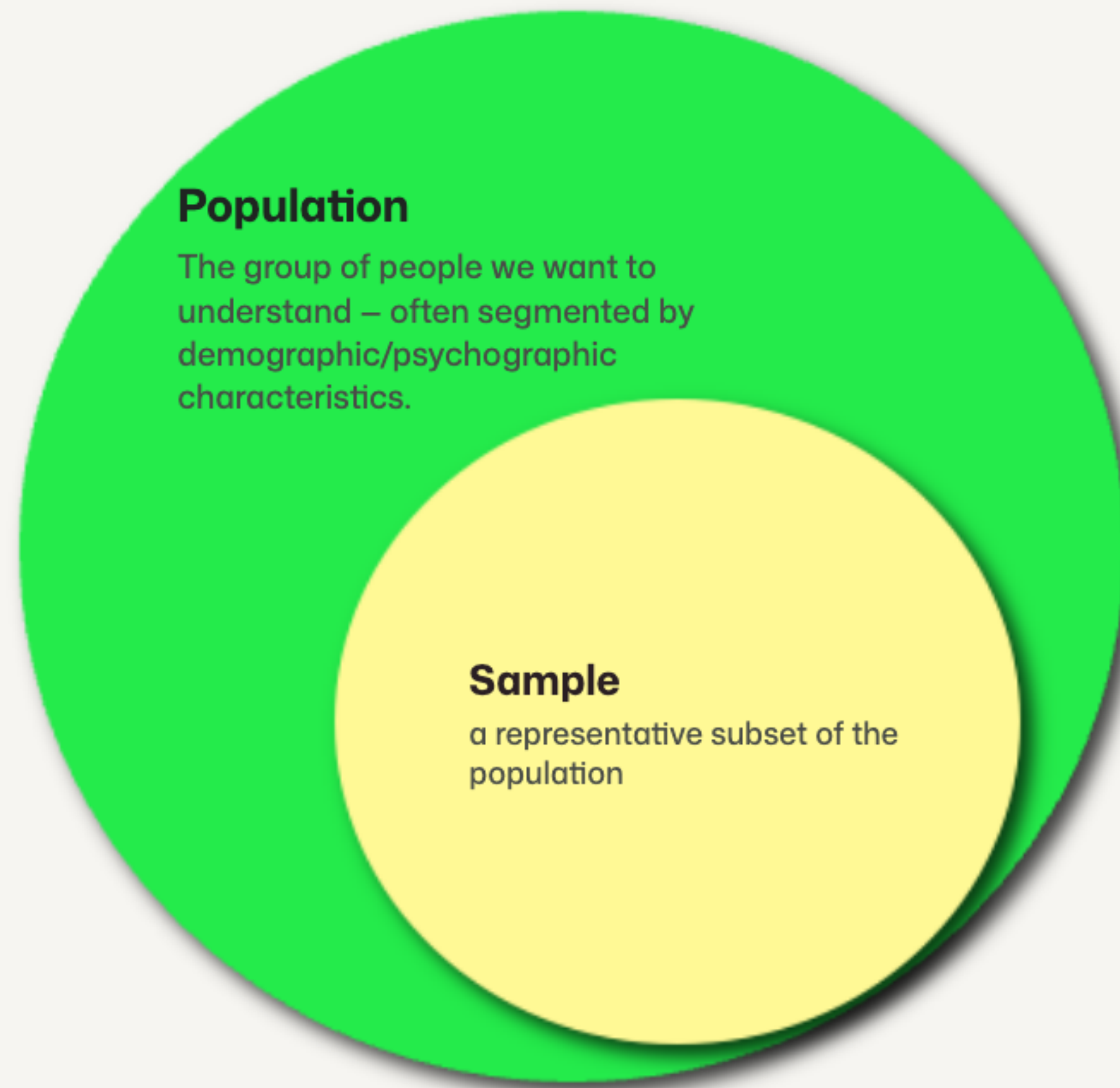
Dewey Defeats Truman

1948: The **Chicago Daily Tribune** announces the wrong election result. President **Harry Truman** beats Thomas Dewey – against all polls.

Reason: a biased, inaccurate opinion poll.



Sampling

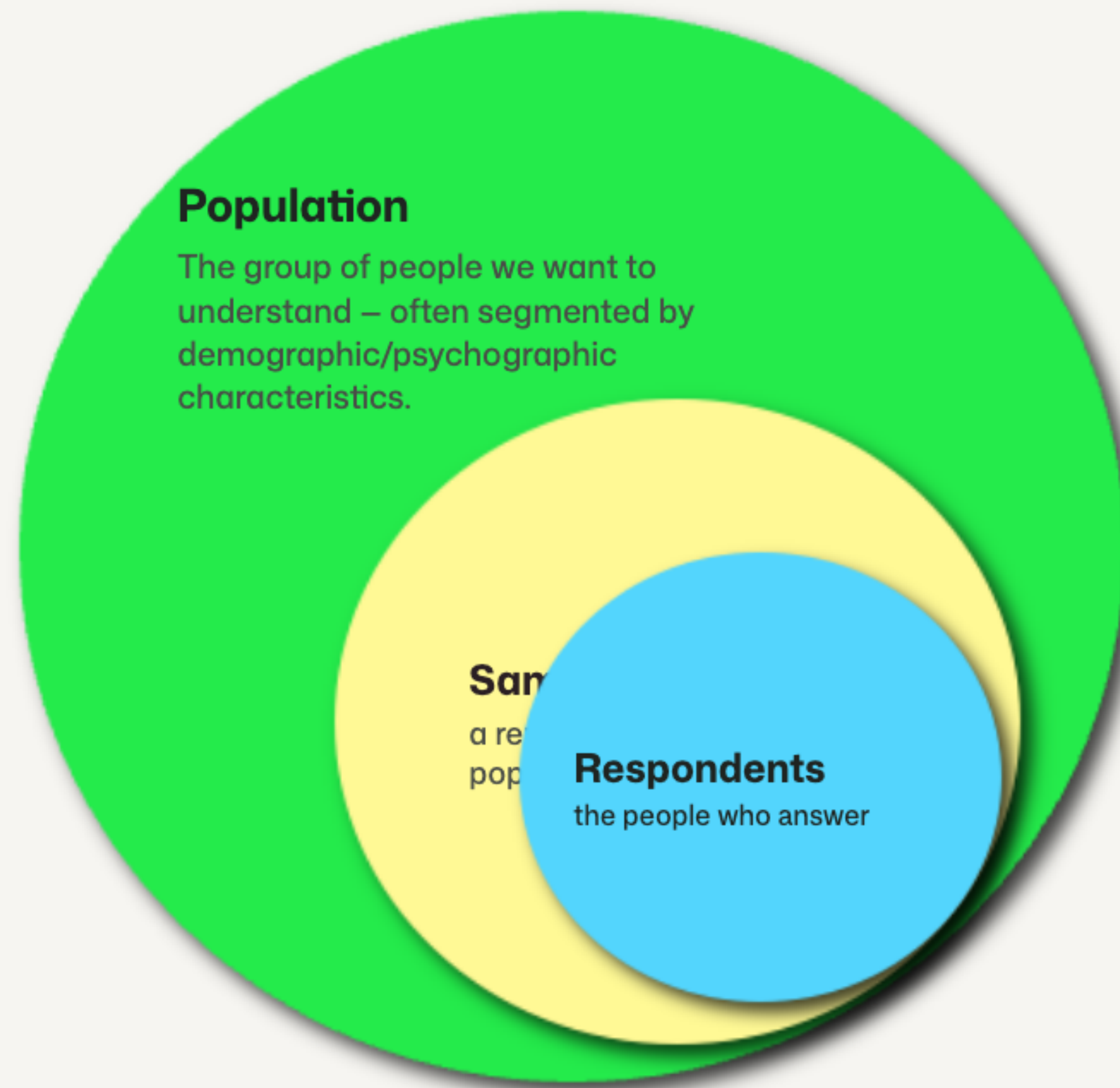


Most surveys **cannot survey every person**. Instead, a **sample** is drawn and examined – this procedure is called **sampling**.

If sampling is done correctly, the survey results can be generalized to the entire **population**.

If the sample is drawn **incorrectly**, **all the data is useless**.

Sampling



But not everyone selected actually answers: those who really take part are the **respondents**.

The respondents are a **subset of the sample** – and only they ultimately provide the data. If this group is skewed, the result is skewed.

Sampling: Two General Methods

METHOD 1

Non-probability Sampling

The sample is drawn based on the **personal judgment of the researcher** – often at random (convenience sample, e.g. passersby in a shopping mall).

Usually **inexpensive**; allows a rough estimate of the population parameters.

But: the **sampling error** cannot be calculated → results are **not representative** and cannot be generalized to the population.

METHOD 2

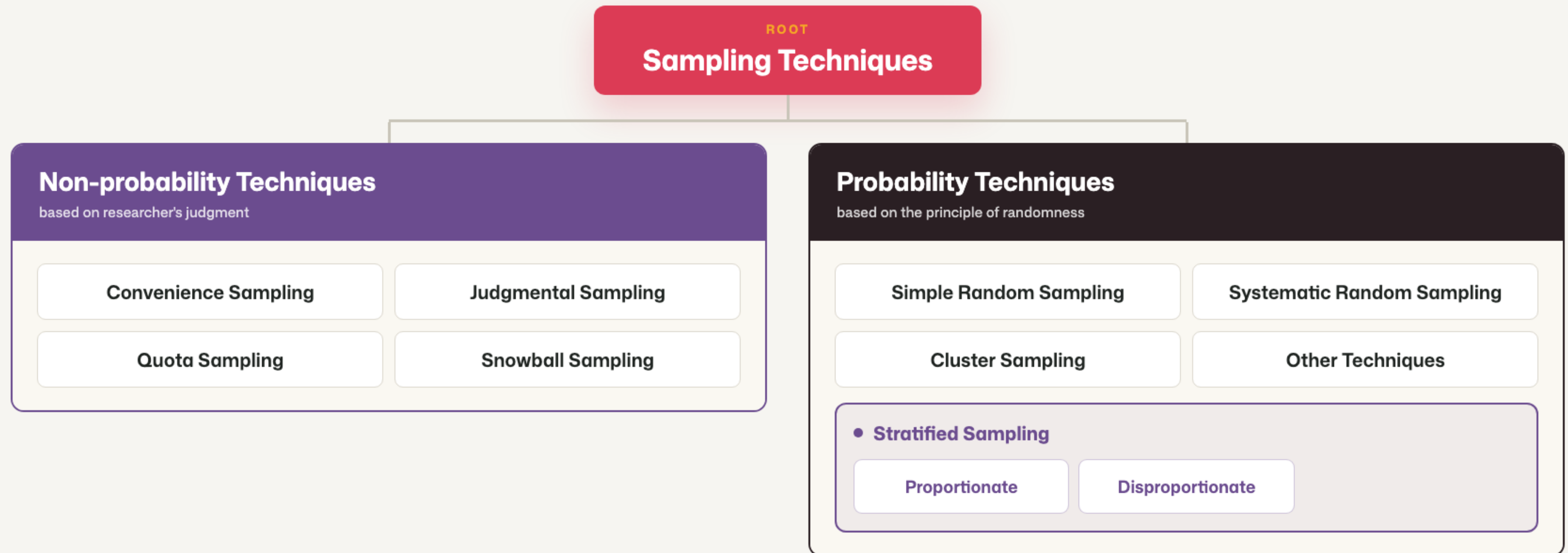
Probability Sampling

The sample is selected based on the **principle of randomness**.

Allows **statistical techniques** to determine the accuracy of the estimated population parameters as well as to assess their **confidence intervals**.

Results are **generalizable** and can be extended to the population.

Sampling Techniques



4

Sampling

4.1 Non-probability Sampling

4.2 Probability Sampling

4.3 Choosing Non-probability vs. Probability Sampling

4.4 Sample Size

Convenience Sampling



In **convenience sampling** (selection at random), respondents enter the sample **uncontrolled** – mostly out of **convenience**. Often simply because they are in the right place at the right time.

TYPICAL EXAMPLES

Students & members of public organizations

Surveys in stores without qualifying the respondents

Surveys on the streets

Tear-off questionnaires in catalogs and magazines

Judgmental Sampling



Judgmental sampling **is a form of convenience sampling in which respondents enter the sample** at the discretion of the researcher..

TYPICAL EXAMPLES

Test markets

Purchasing engineers in industrial market research

Mothers as "users" of diapers

Quota Sampling

The sample is drawn according to **predefined control characteristics** (e.g. gender, age, income), so that it **reflects the structure of the population** proportionally. The objects are usually selected at random – but they must fulfill the quota plan.

CONTROL CHARACTERISTIC	POPULATION	SAMPLE	
	SHARE %	SHARE %	COUNT
Gender – male	48	48	480
female	52	52	520
Total	100	100	1000
Age – 18–30	27	27	270
31–45	39	39	390
45–60	16	16	160
over 60	18	18	180
Total	100	100	1000

Often used in online surveys.

Snowball Sampling

also chain sampling



- 1 The first group of respondents is (usually) selected **randomly**.
- 2 After the interview, they name **additional people** from the target group.
- 3 Subsequent respondents are selected through **referrals**.

Good for **locating rare characteristics**: hard-to-reach respondents (public officials, executives, homeless people, drug addicts), rarely occurring traits, buyer-seller pairs in industrial research.

4

Sampling

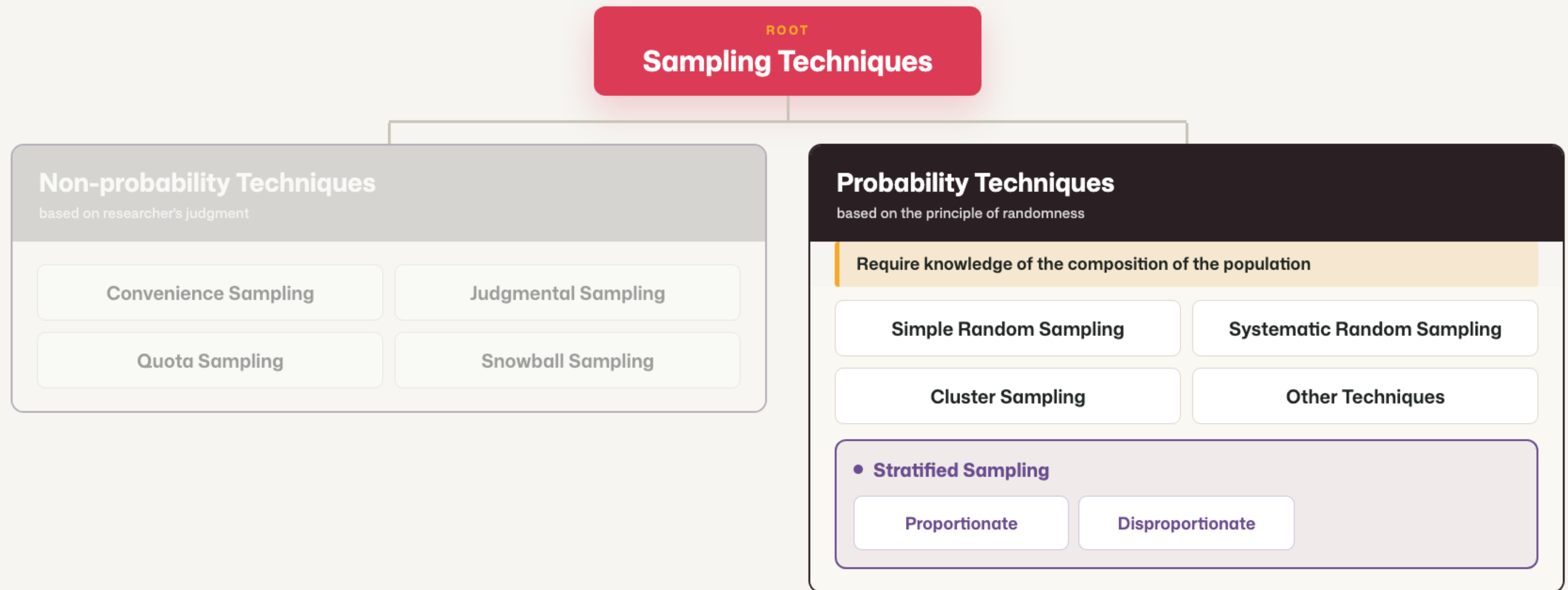
4.1 Non-probability Sampling

4.2 Probability Sampling

4.3 Choosing Non-probability vs. Probability Sampling

4.4 Sample Size

Sampling Techniques



Simple and Systematic Random Sampling

Simple Random Sampling

- Each element is selected independently of all others. This means:
- Each element of the population has a known and equal probability of being selected.
- Every possible sample of size n has a known probability of actually being drawn.

SELECT AT
RANDOM

Abolina	Tirza
Bernhardt	Carina
● Berz	Helena
Boeck	Nicola
Dollase	Miriam
Fränzel	Carolin
Frost	Annika
Goetze	Anika
● Jähel	Felix
Jank	Nadja
Keitzl	Inga
● Kropp	Janine
Kubitzky	Victoria
Langen	Eduard

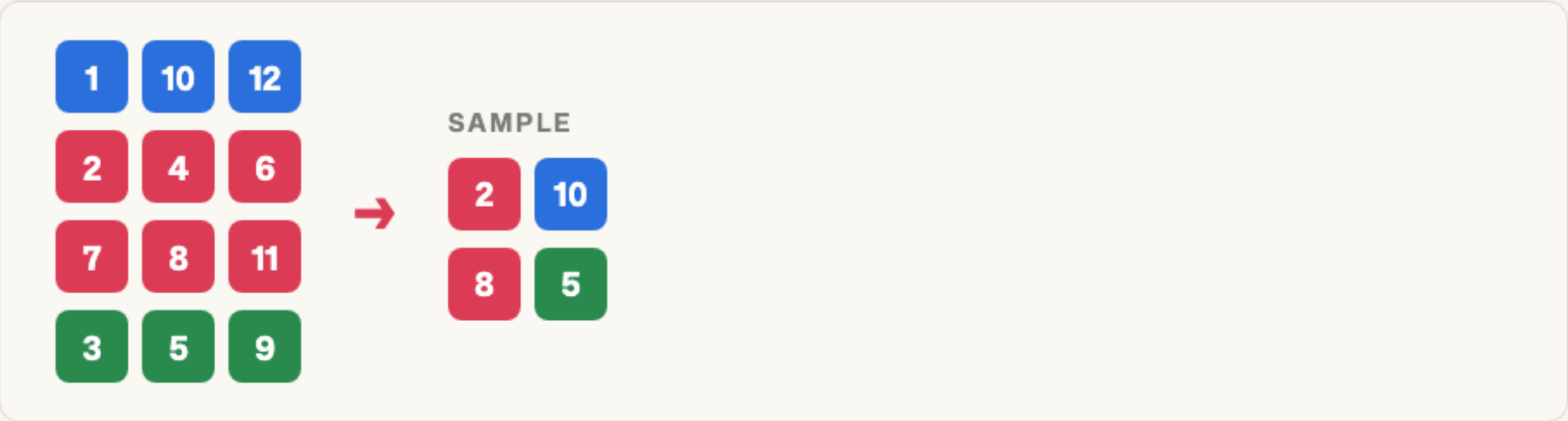
Systematic Random Sampling

- First, a starting element is selected at random; then every i-th element is drawn from the sampling frame.
- The interval i results from the size of the population N relative to the size of the sample n: $i = N / n$

START
HERE, THEN
EVERY
I
-TH

● Abolina	Tirza
i Bernhardt	Carina
Berz	Helena
● Boeck	Nicola
i Dollase	Miriam
Fränzel	Carolin
● Frost	Annika
i Goetze	Anika
Groth	Carolin
● Jähel	Felix
i Jank	Nadja
Keitzl	Inga
● Kropp	Janine
Kubitzky	Victoria

Stratified Sampling



The population is first divided into **non-overlapping strata**. Then a **(dis-)proportional** share is drawn at random from each stratum. Elements within a stratum should be **similar** to one another.

GOOD FOR

- Highlighting a particular subgroup within the population
- Observing relationships between two or more subgroups
- Representative sampling of even the smallest and least accessible subgroups
- Higher statistical precision

Proportional

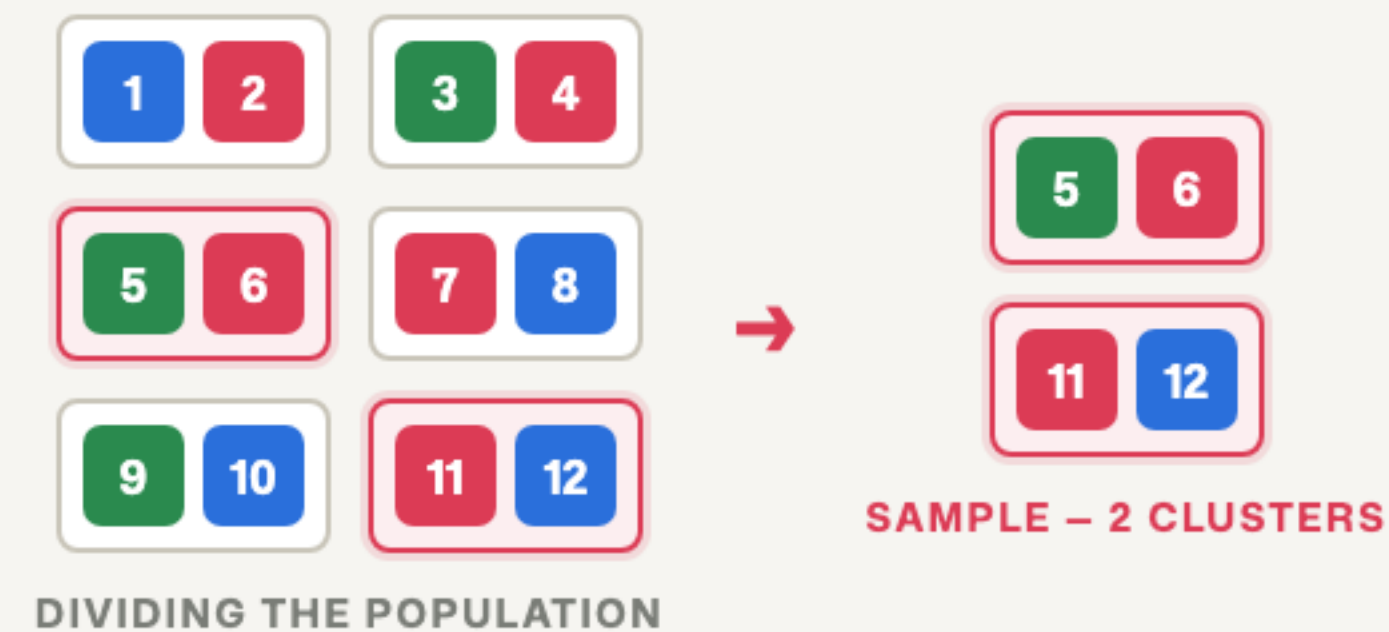
STRATUM	A	B	C
Population size	100	200	300
Sampling fraction	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Sample size	50	100	150

Disproportional

STRATUM	A	B	C
Population size	100	200	300
Sampling fraction	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$
Sample size	20	100	100

Cluster Sampling

also called cluster sampling



The population is divided into **exclusive clusters**. Then **entire clusters are selected at random** and enter the sample in full.

For each cluster, either **all elements** (one-stage) or a **random subsample** (two-stage) is drawn.

GOOD FOR

Covering large geographic areas

Reducing (survey) costs

When a complete list of elements is difficult to compile

When the population consists of natural clusters (blocks, cities, schools, hospitals ...)

4

Sampling

4.1 Non-probability Sampling

4.2 Probability Sampling

4.3 Choosing Non-probability vs. Probability Sampling

4.4 Sample Size

Strengths and Weaknesses of Basic Sampling Techniques

TECHNIQUE	STRENGTHS	WEAKNESSES
NON-PROBABILITY SAMPLING TECHNIQUES		
Convenience Sampling	+ Least expensive, least time-consuming, most convenient	- Prone to error, not representative; not recommended for descriptive and causal research
Judgmental Sampling	+ Low cost, convenient, not time-consuming	- Subjective, results not generalizable
Quota Sampling	+ Certain characteristics of the sample can be controlled	- Prone to error, no guarantee of representativeness
Snowball Sampling	+ Enables estimation of rare characteristics	- Time-consuming in fieldwork
PROBABILITY SAMPLING TECHNIQUES		
Simple Random Sampling	+ Easy to understand; generalizable or representative results	- Sampling frame difficult to construct, expensive, lower precision; no guarantee of representativeness
Systematic Sampling	+ Can increase representativeness; easier to implement than simple random sampling	- Can decrease representativeness
Stratified Sampling	+ Includes all important subgroups of the population; high precision	- Relevant stratification criteria difficult to select; multiple criteria not practical; expensive
Cluster Sampling	+ Easy to implement, cost-effective	- Imprecise; complicated computation and interpretation of results

4

Sampling

4.1 Non-probability Sampling

4.2 Probability Sampling

4.3 Choosing Non-probability vs. Probability Sampling

4.4 Sample Size

Determining the Sample Size

The sample size does **not** depend on the size of the population – it is determined by the **qualitative aspects** of the study:

1

Desired precision of the predictions

5

Importance of the decision

2

Knowledge about the population parameters

6

Response and drop-out rates

3

Number of variables

7

Resource constraints

4

Type of analysis

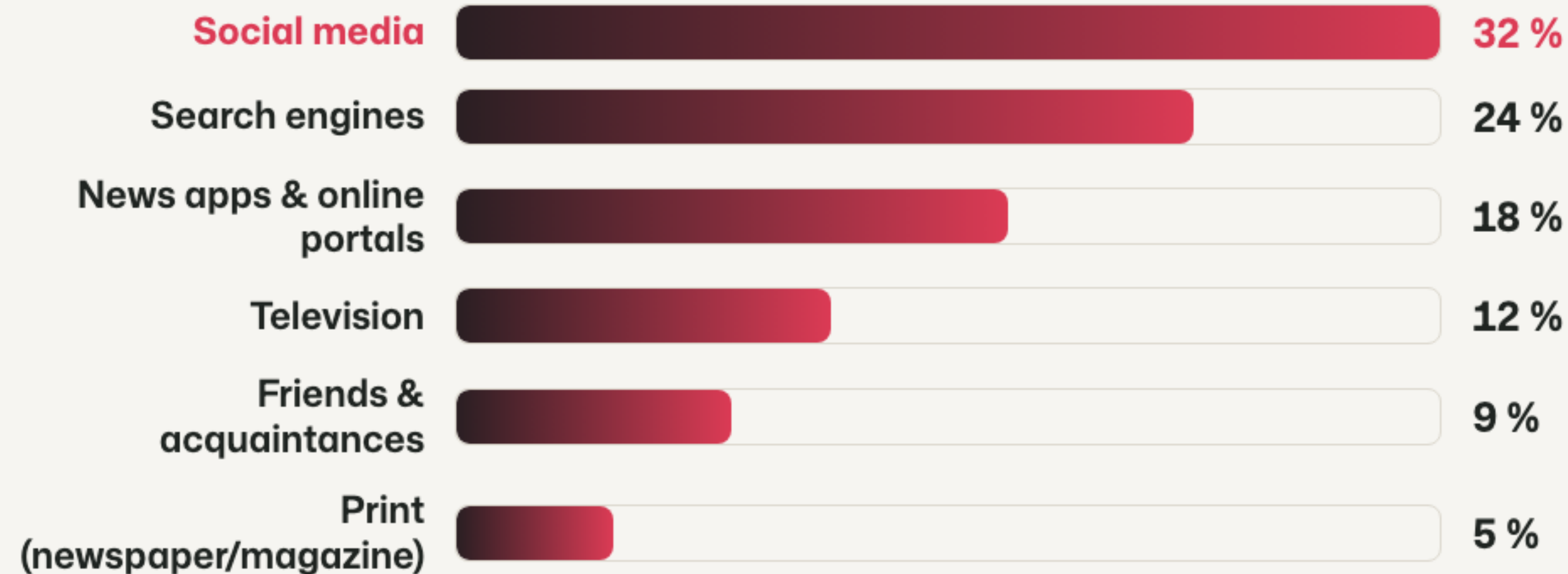


Sample Sizes Used in Marketing Research Studies

TYPE OF STUDY	MINIMUM SIZE	TYPICAL SIZE
Problem identification studies (e.g. market potential)	500	1.000 – 2.000
Problem-solving studies (e.g. pricing)	200	300 – 500
Product tests	200	300 – 500
Test market studies	200	300 – 500
TV/radio/print advertising (per ad)	150	200 – 300
Test market audits	10 stores	10 – 20 stores
Focus groups	6 groups	10 – 15 groups

A survey result – and how certain is it?

“Which sources do you prefer for obtaining your information?” (n = 48,804)



32 %

name **social media** as their main information channel.

But how close is this value to the true value in the population?

The answer is given by the **margin of error** – and it determines the required sample size.

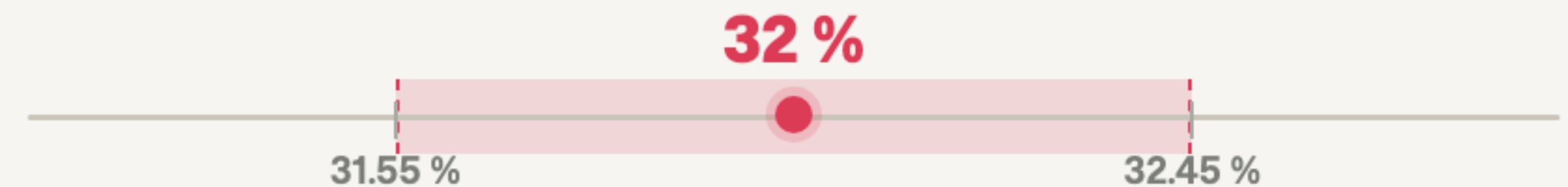
Margin of Error Approach to Determining Sample Size

MARGIN OF ERROR

Margin of error is the measure of a survey's precision.

- The **smaller** the margin of error, the **more precise** the survey's estimates are.

RESULT WITH MARGIN OF ERROR



Margin of error approach: two formulas

$$x = \hat{x} \pm E$$

x = true value of the parameter $\hat{x}$ = sample value E = margin of error

FOR METRIC DATA

Means

Assessing means computed from the sample

$$E = z \frac{\sigma}{\sqrt{n}}$$

- z z-value for the specified confidence level
- σ Standard deviation of the parameter in the population
- n Sample size

FOR PROPORTIONS

Proportions

Assessing computed proportions

$$E = z \cdot \sqrt{\frac{\pi(1 - \pi)}{n}}$$

- z z-value for the specified confidence level
- π Estimated value for the proportion in the population
- n Sample size

Margin of Error Approach: Two Formulas

$$x = \hat{x} \pm E$$

x = true value of the parameter $\hat{x}$ = sample value E = margin of error

FOR METRIC DATA

Means

Assessing means calculated from the sample

$$E = z \frac{\sigma}{\sqrt{n}}$$

Usually unknown

- z z-value for the specified confidence level
- σ Standard deviation of the parameter in the population
- n Sample size

FOR PROPORTIONS

Proportions

Assessing calculated proportions

$$E = z \cdot \sqrt{\frac{\pi(1 - \pi)}{n}}$$

Usually unknown

- z z-value for the specified confidence level
- π Estimate of the proportion in the population
- n Sample size

Margin of Error Approach: Two Formulas

$$x = \hat{x} \pm E \quad \text{\textcolor{red}{x}} = \text{true value of the parameter} \quad \text{\textcolor{red}{\hat{x}}} = \text{sample value} \quad \text{\textcolor{red}{E}} = \text{margin of error}$$

FOR METRIC DATA

Means

Assessing means calculated from the sample

$$E = z \frac{\sigma}{\sqrt{n}}$$

Usually unknown

- z** z-value for the specified confidence level
- σ** Standard deviation of the parameter in the population
- n** Sample size

FOR PROPORTIONS

Proportions

Assessing calculated proportions

$$E = z \sqrt{\frac{\pi(1 - \pi)}{n}}$$

Usually unknown

Maximum at $\pi = 0.5$

- z** z-value for the specified confidence level
- π** Estimate of the proportion in the population
- n** Sample size

z-Values and Maximum Margin of Error

Z-VALUES

1.96 for **95 %** confidence level

2.58 for **99 %** confidence level

MAXIMUM MARGIN OF ERROR FOR 95% CONFIDENCE LEVEL

With **z = 1.96** and the maximum **$\pi = 0.5$** the formula simplifies to:

$$\begin{aligned} E &= z \cdot \sqrt{\frac{\pi(1 - \pi)}{n}} \\ &= 1,96 \cdot \sqrt{\frac{0,5(1 - 0,5)}{n}} \\ &\approx \frac{1}{\sqrt{n}} \end{aligned}$$

... the upper bound of the error – independent of the actual proportion π .

What Is the Margin of Error?

"Which sources do you prefer to get your information from?"

32 % name **social media** as their main information channel

SAMPLE **48.804** Respondents

$$\text{Margin of error} = \frac{1}{\sqrt{n}}$$

48.804 respondents in the sample

$$\sqrt{48.804} = 220,916 \approx 221$$

$$1 / 221 = 0.0045$$

$$\cdot 100 = 0.45\%$$

$$\Rightarrow x = 32\% \pm 0.45\%$$

$$\Rightarrow \text{from } 31.55\% \text{ to } 32.45\%$$

Calculations show approximate values for a 95% confidence level

How Large Must the Sample Be?

$$E \approx \frac{1}{\sqrt{n}} \Rightarrow n \approx \left(\frac{1}{E}\right)^2$$

The higher the desired accuracy, the larger the sample must be.

± 1 %

$$n = (1 / 0.01)^2 = 100^2$$

10.000

± 2 %

$$n = (1 / 0.02)^2 = 50^2$$

2.500

± 5 %

$$n = (1 / 0.05)^2 = 20^2$$

400

± 10 %

$$n = (1 / 0.1)^2 = 10^2$$

100

Doubling the accuracy costs four times the sample. The size does **not** depend on the size of the population.

Calculations show approximate values for a 95% confidence level

What If the Population Is Small?

For a margin of error of $\pm 1\%$ you would need:

$$n = (1 / 0.01)^2 = 100^2$$

10.000

But what if the population comprises only **100 elements?** (e.g. car manufacturers)

RULE OF THUMB

If the sample is **larger than 10%** of the population, **corrections** are necessary.

Otherwise the formula overestimates the required size – the finite population reduces the actual sampling error.

Calculations show approximate values for a 95% confidence level

Correcting the Sample Size

$$n_{\text{corr}} = \frac{n \cdot N}{n + N - 1}$$

n_{corr} corrected sample size

n (uncorrected) sample size

N Population size

EXAMPLE · $N = 100$

Computationally you would need $n = 10,000$ – with only **100** elements in the population:

$$\begin{aligned} n_{\text{corr}} &= (10.000 \cdot 100) / (10.000 + 100 - 1) \\ &= 1.000.000 / 10.099 \approx 99 \end{aligned}$$

You can't survey more than the entire population – the correction brings the size down to a realistic level.

Calculations show approximate values for a 95% confidence level

Correction for a small population: $\pm 1\%$

$$n \approx \left(\frac{1}{E}\right)^2$$

$$n_{\text{corr}} = \frac{n \cdot N}{n + N - 1}$$

COMPUTATIONALLY REQUIRED

$$n = (1 / 0.01)^2 = 100^2$$

10.000

... with only **N = 100** elements in the population.



CORRECTED

$$n_{\text{corr}} = (10.000 \cdot 100) / (10.000 + 100 - 1) \\ = 1.000.000 / 10.099$$

≈ 99

Calculations show approximate values for a 95% confidence level

Correction for a small population: $\pm 5\%$

$$n \approx \left(\frac{1}{E}\right)^2$$

$$n_{\text{corr}} = \frac{n \cdot N}{n + N - 1}$$

COMPUTATIONALLY REQUIRED

$$n = (1 / 0.05)^2 = 20^2$$

400

... with only **N = 100** elements in the population.



CORRECTED

$$n_{\text{corr}} = (400 \cdot 100) / (400 + 100 - 1) \\ = 40.000 / 499$$

≈ 80

Calculations show approximate values for a 95% confidence level

Correction for a small population: $\pm 10\%$

$$n \approx \left(\frac{1}{E}\right)^2$$

$$n_{\text{corr}} = \frac{n \cdot N}{n + N - 1}$$

COMPUTATIONALLY REQUIRED

$$n = (1 / 0.1)^2 = 10^2$$

100

... with only **N = 100** elements in the population.



CORRECTED

$$n_{\text{corr}} = (100 \cdot 100) / (100 + 100 - 1) \\ = 10.000 / 199$$

≈ 50

Calculations show approximate values for a 95% confidence level

Confidence Interval and Confidence Level

CONFIDENCE INTERVAL

An estimated **range of values** together with the probability that this range contains the unknown parameter value.

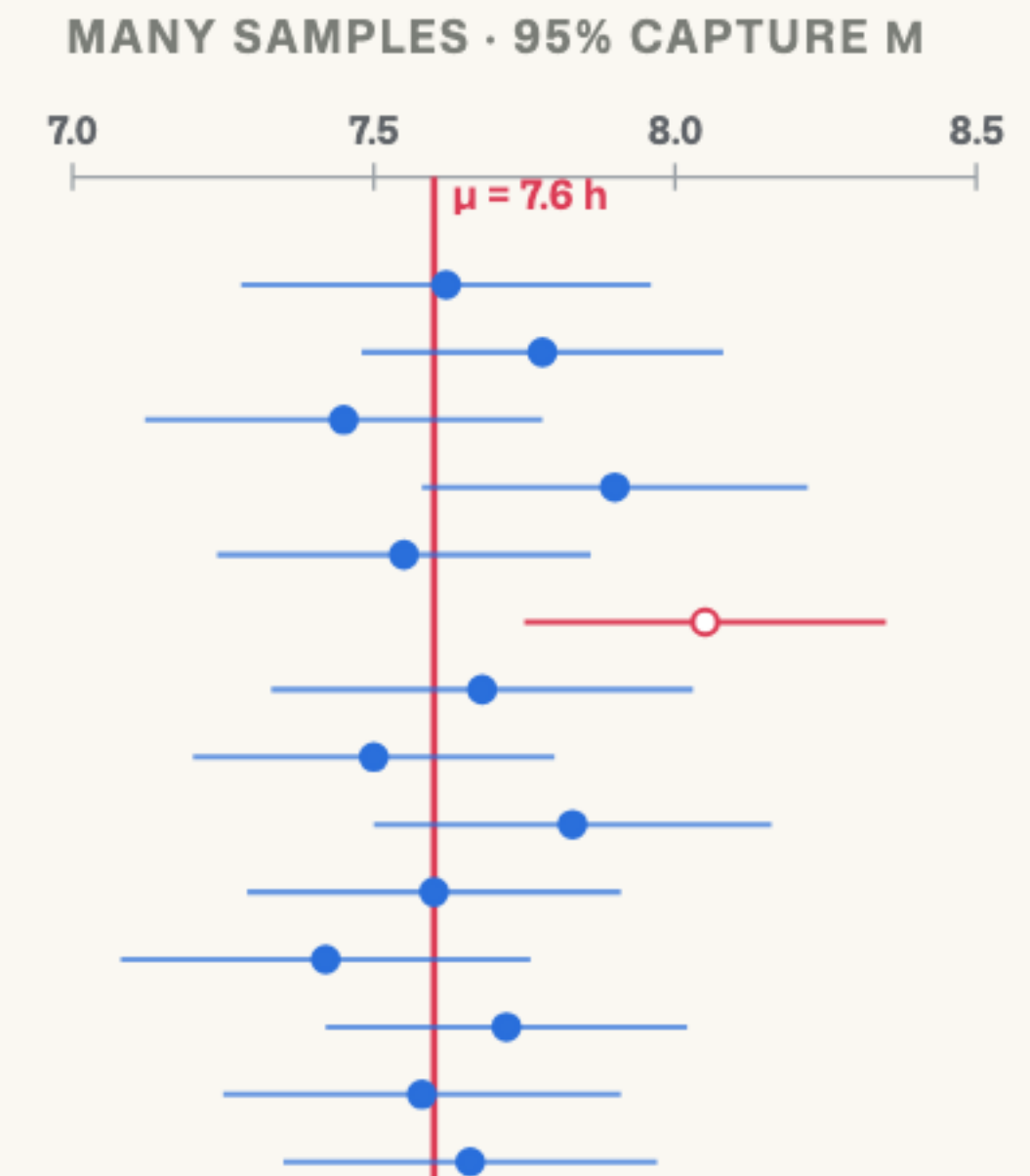
CONFIDENCE LEVEL

The **expected proportion** of intervals that contain the parameter value across many samples.

EXAMPLE · WORKING HOURS

Sample of 30 people → avg. **7.5 h**. Confidence interval: **7.2 – 7.8 h** (margin of error ± 0.3).

95% confidence level means: if you repeated the measurement **100 times** with new samples, the true average would fall within this range in **95** of the cases.



Confidence Interval, Margin of Error, and Sample Size

The higher the **certainty** (confidence probability) we need, the **wider** the confidence interval becomes – and the **larger** the margin of error.

1.96 for 95%
→ margin of error $\approx 1 / \sqrt{n}$

2.58 for 99%
→ margin of error $\approx 1.29 / \sqrt{n}$

More certainty → larger z-value → for the same precision you need a **larger sample**.

$$\begin{aligned}
 & \text{95 \%} \\
 E &= z \cdot \sqrt{\frac{\pi(1 - \pi)}{n}} \\
 &= 1,96 \cdot \sqrt{\frac{0,5(1 - 0,5)}{n}} \\
 &\approx \frac{1}{\sqrt{n}}
 \end{aligned}$$

$$\begin{aligned}
 & \text{99 \%} \\
 E &= z \cdot \sqrt{\frac{\pi(1 - \pi)}{n}} \\
 &= 2,58 \cdot \sqrt{\frac{0,5(1 - 0,5)}{n}} \\
 &= \frac{1,29}{\sqrt{n}}
 \end{aligned}$$

What the Margin of Error Formula Tells Us

$$E = z \cdot \sqrt{\frac{\pi(1 - \pi)}{n}}$$

n ↑
lowers E

z ↑
raises E

- 1 The only lever to **lower** the margin of error is a larger **n** – everything else is fixed.
- 2 More certainty means a larger **z** → the margin of error **grows** → you need an **even larger n** to push it back down.

Smaller margins of error require larger samples.

Higher confidence levels require larger samples.

5

Data Analysis: A Concise Overview of Statistical Techniques

5.1 Descriptive Statistics: Organizing and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

Types of Statistical Data Analysis

TYPE 1

Descriptive Statistics

Summarizes the observations from the sample and presents them **clearly**.

Uses **summary measures, tables, graphs and charts** to describe, systematize, organize and present the collected data.

TYPE 2

Inferential Statistics

Makes statements about the **generalizability** of observations and conclusions from random samples to the population.

Assesses **relationships between variables** and quantifies them: strength and significance, predictions and estimates.

5

Data Analysis

5.1 Descriptive Statistics: Organizing and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

5

Data Analysis

5.1 Descriptive Statistics: Organizing and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

Frequencies and Relative Frequencies

COLLECTED DATA · FAVORITE COLOR (N = 26)

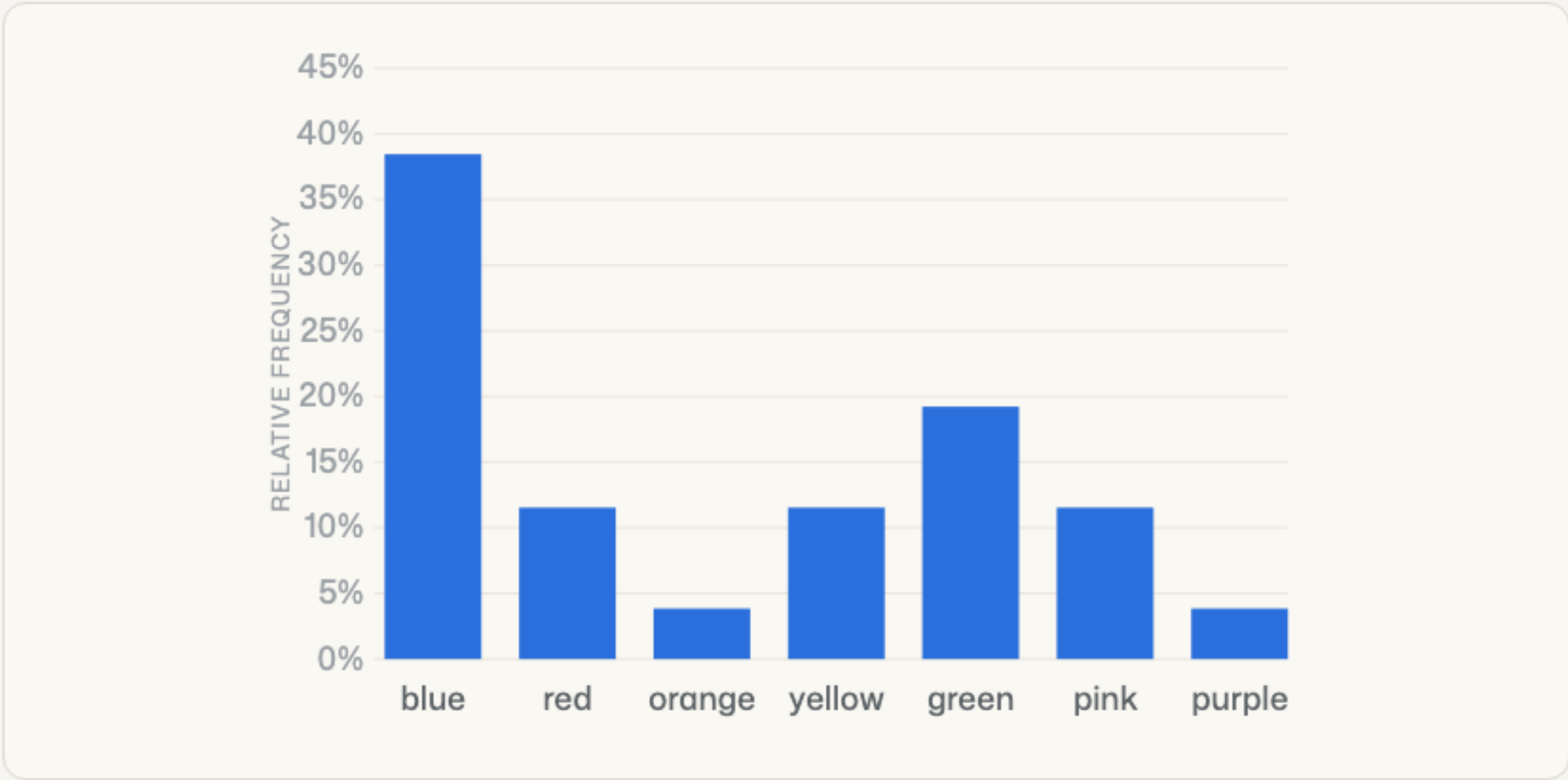
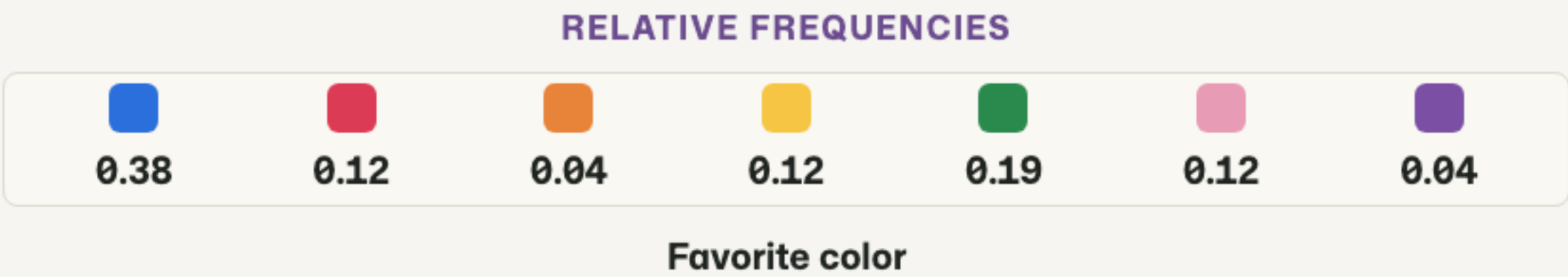
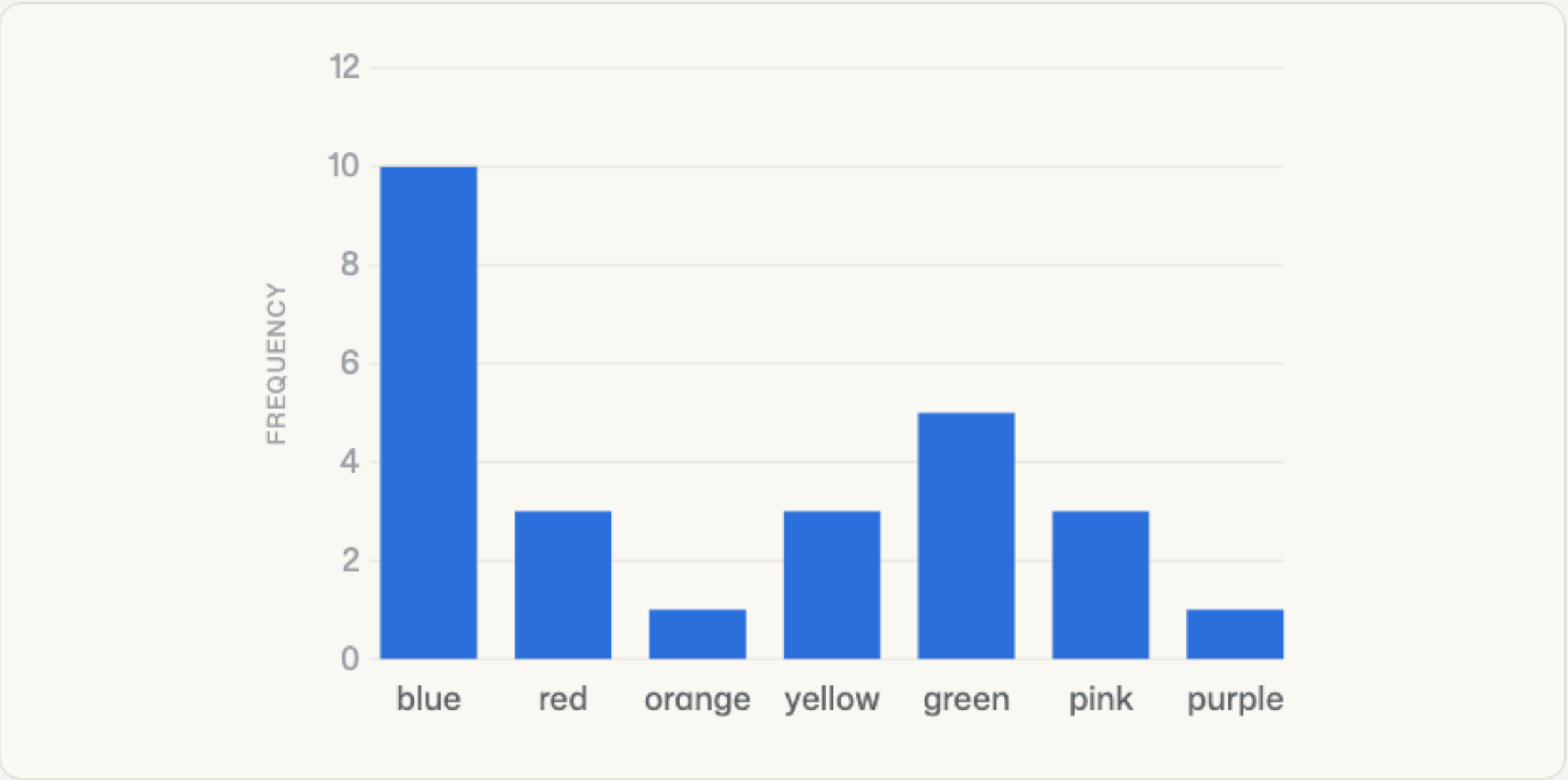
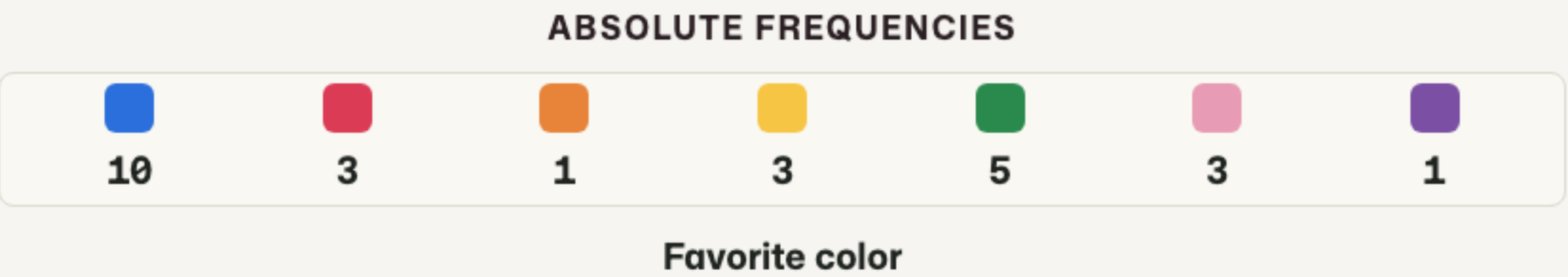


- **Frequency distribution** indicates, for each value, how often it occurs in the data.
- **Relative frequency** shows the proportion (or percentage) of observations for a value.

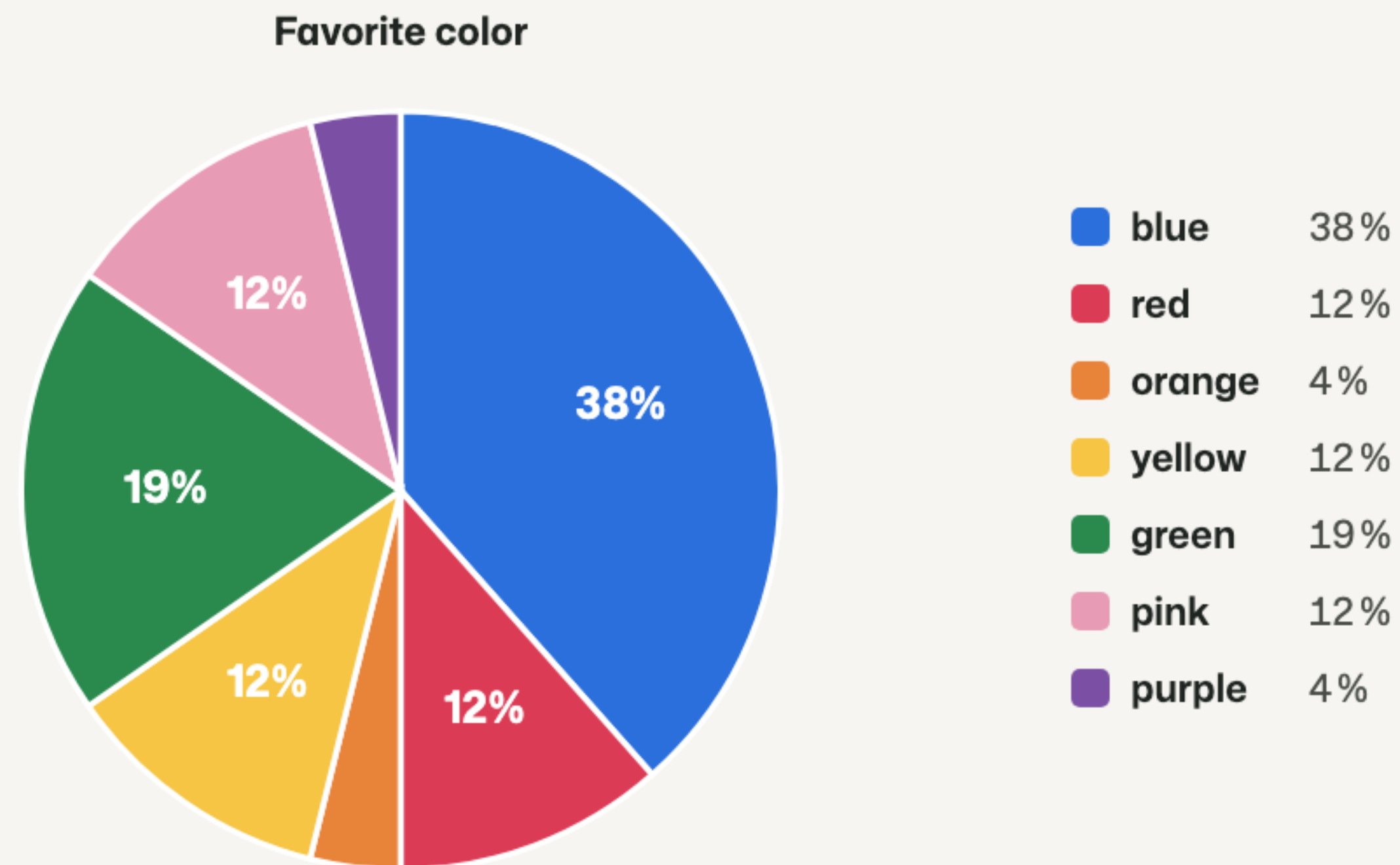
FAVORITE COLOR	FREQUENCY	RELATIVE FREQUENCY
blue	10	$10/26 \approx 0.38$
red	3	$3/26 \approx 0.12$
orange	1	$1/26 \approx 0.04$
yellow	3	$3/26 \approx 0.12$
green	5	$5/26 \approx 0.19$
pink	3	$3/26 \approx 0.12$
purple	1	$1/26 \approx 0.04$
Total	26	1.00

Bar Graph

- Bar height = **frequency** or **relative frequency**
- Bars must **not touch**



Pie Chart



- Should always show **relative frequencies**.
- Needs **labels** – directly on the chart or in the legend.

5

Data Analysis

5.1 Descriptive Statistics: Organizing and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

Tables

Collected Data

2	2	2	4	5	3	3	3	3	2	1	2	3	5
3	4	3	1	2	3	5	3	2	1	3	2		

NUMBER OF CHILDREN	FREQUENCY	RELATIVE FREQUENCY
1	3	$3/26 \approx 0.12$
2	8	$8/26 \approx 0.31$
3	10	$10/26 \approx 0.38$
4	2	$2/26 \approx 0.08$
5	3	$3/26 \approx 0.12$

Discrete Variable is a quantitative variable that has either a finite number of values or an infinitely countable number of values (e.g. 0, 1, 2, 3, ...).

Sometimes there are **too many values** to create a row for each value. In that case, several values are combined into **groups (classes)**.

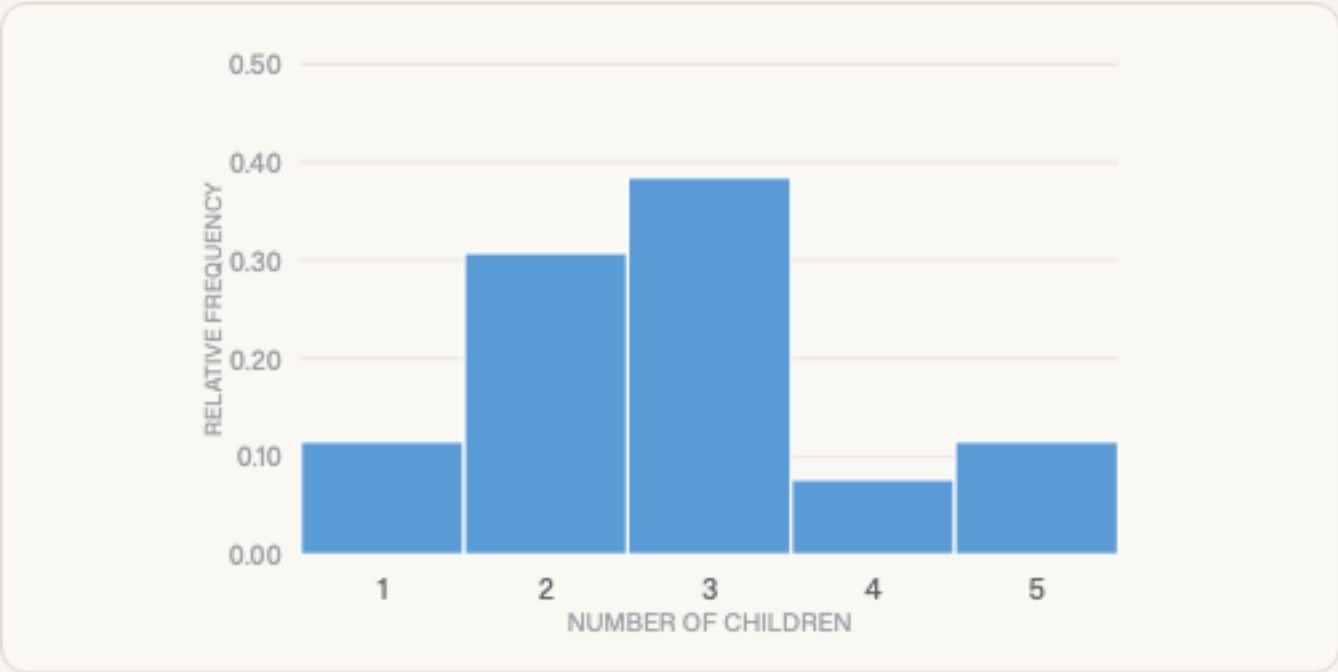
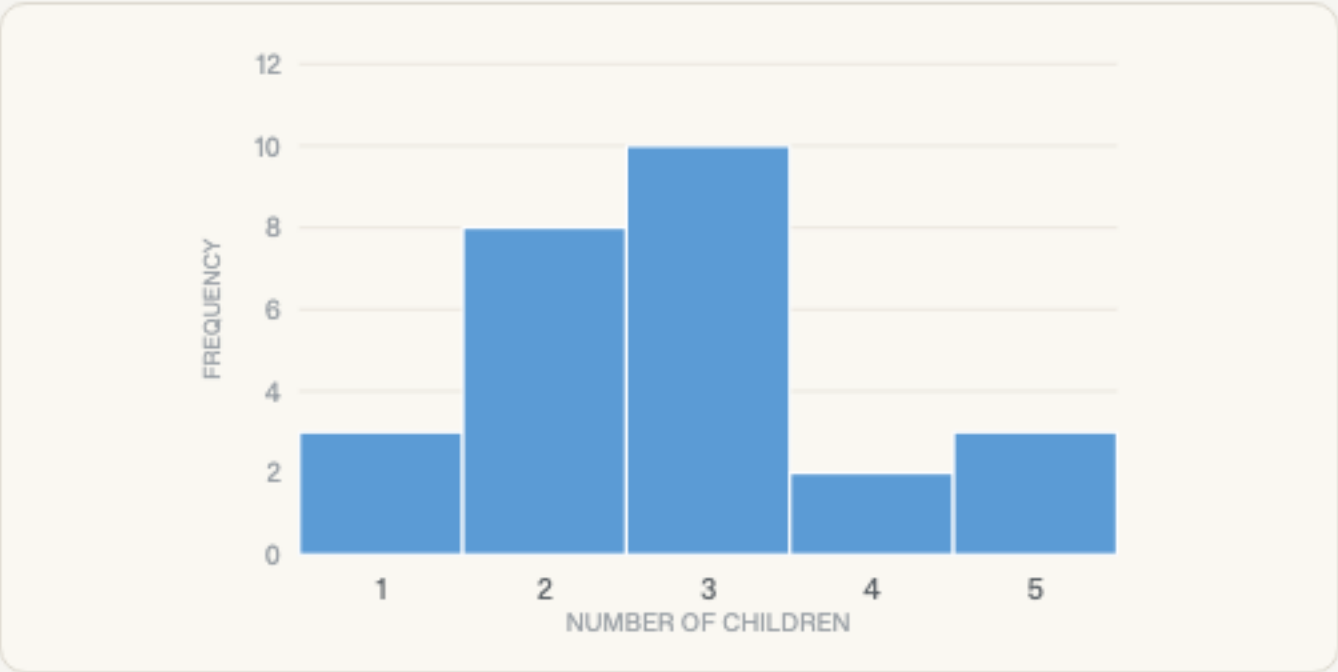
Collected Data

62	87	67	58	95	94	91	69	52	76	82
85	91	60	77	72	83	79	63	88	79	88
70	75	75								

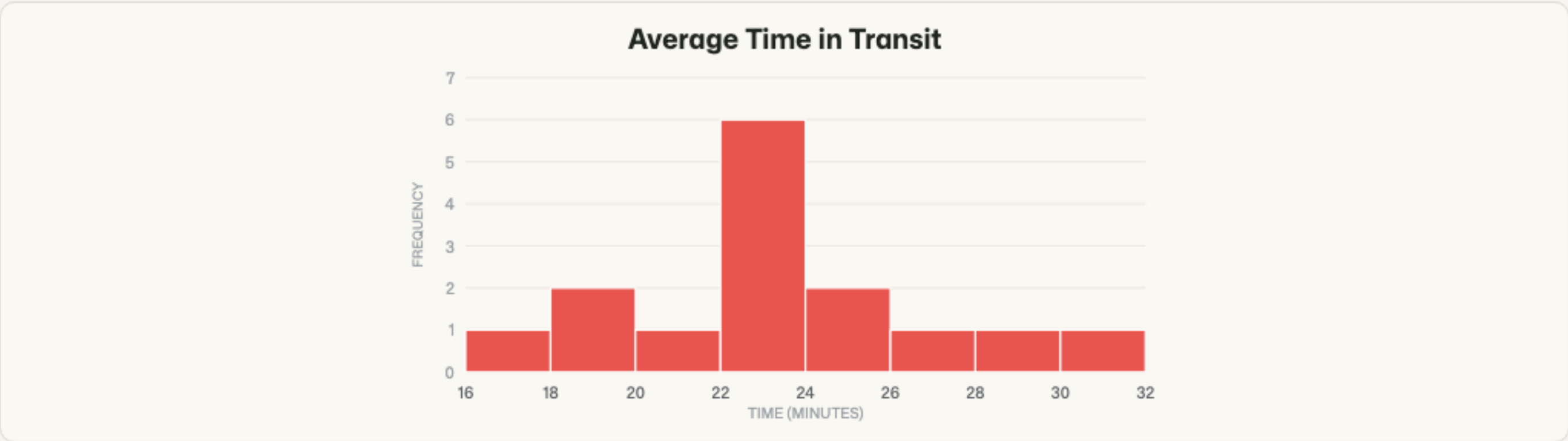
	POINTS ON THE EXAM	FREQUENCY
Lower Class Limit →	50–59	2
Upper Class Limit →	60–69	5
	70–79	7
	80–89	7
Class Width = 90 – 80 = 10 →	90–99	4

Tables and Histograms

NUMBER OF CHILDREN	FREQUENCY	RELATIVE FREQUENCY
1	3	$3/26 \approx 0.12$
2	8	$8/26 \approx 0.31$
3	10	$10/26 \approx 0.38$
4	2	$2/26 \approx 0.08$
5	3	$3/26 \approx 0.12$

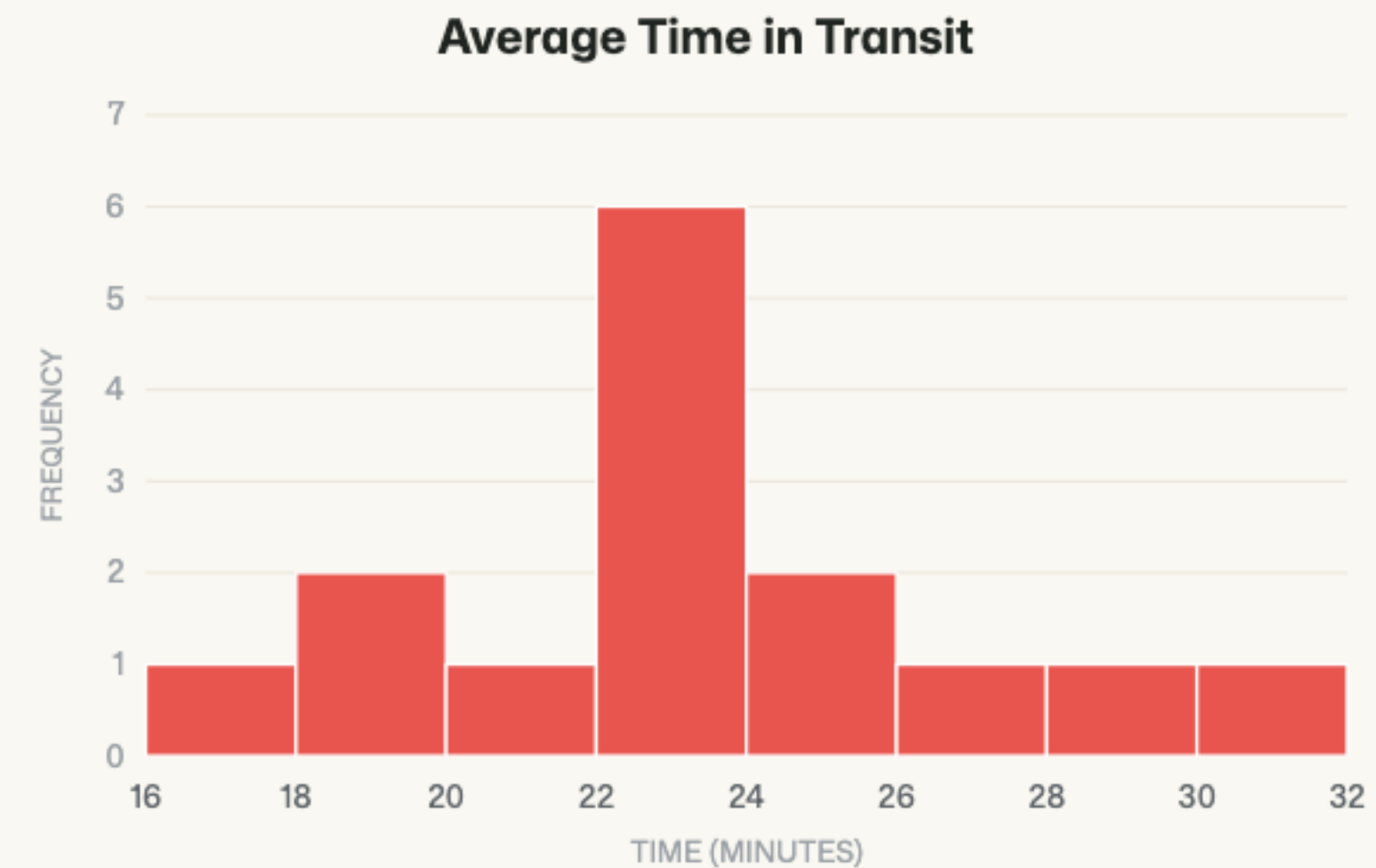


⌚ TIME IN TRANSIT	FREQUENCY	RELATIVE FREQUENCY
16-17.9	1	$1/15 \approx 0.07$
18-19.9	2	$2/15 \approx 0.13$
20-21.9	1	$1/15 \approx 0.07$
22-23.9	6	$6/15 \approx 0.40$
24-25.9	2	$2/15 \approx 0.13$
26-27.9	1	$1/15 \approx 0.07$
28-29.9	1	$1/15 \approx 0.07$
30-31.9	1	$1/15 \approx 0.07$



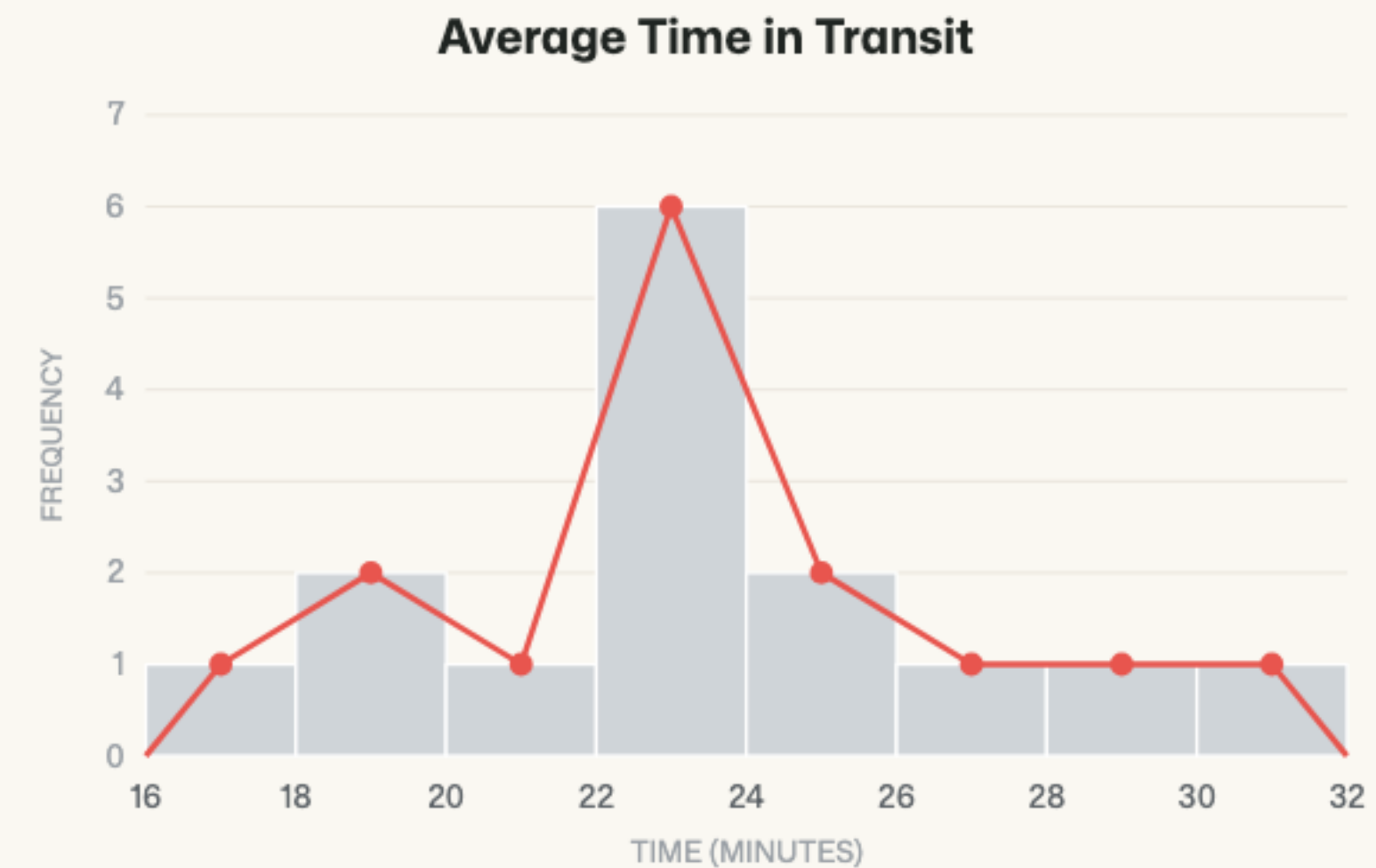
Histogram

- A **histogram** graphically depicts a grouped frequency distribution.
- The **height** of each bar corresponds to the frequency (or relative frequency) of the class.
- The **widths** are equal and the bars **touch each other** – unlike in a bar graph.



Frequency Polygon

- 1 Mark the **midpoint** at the top of each bar of the histogram.
- 2 **Connect** the midpoints with straight lines.
- 3 Bring the line back down to **zero** at both ends – and the polygon is complete.



Cumulative Tables and Ogives

Cumulative Table

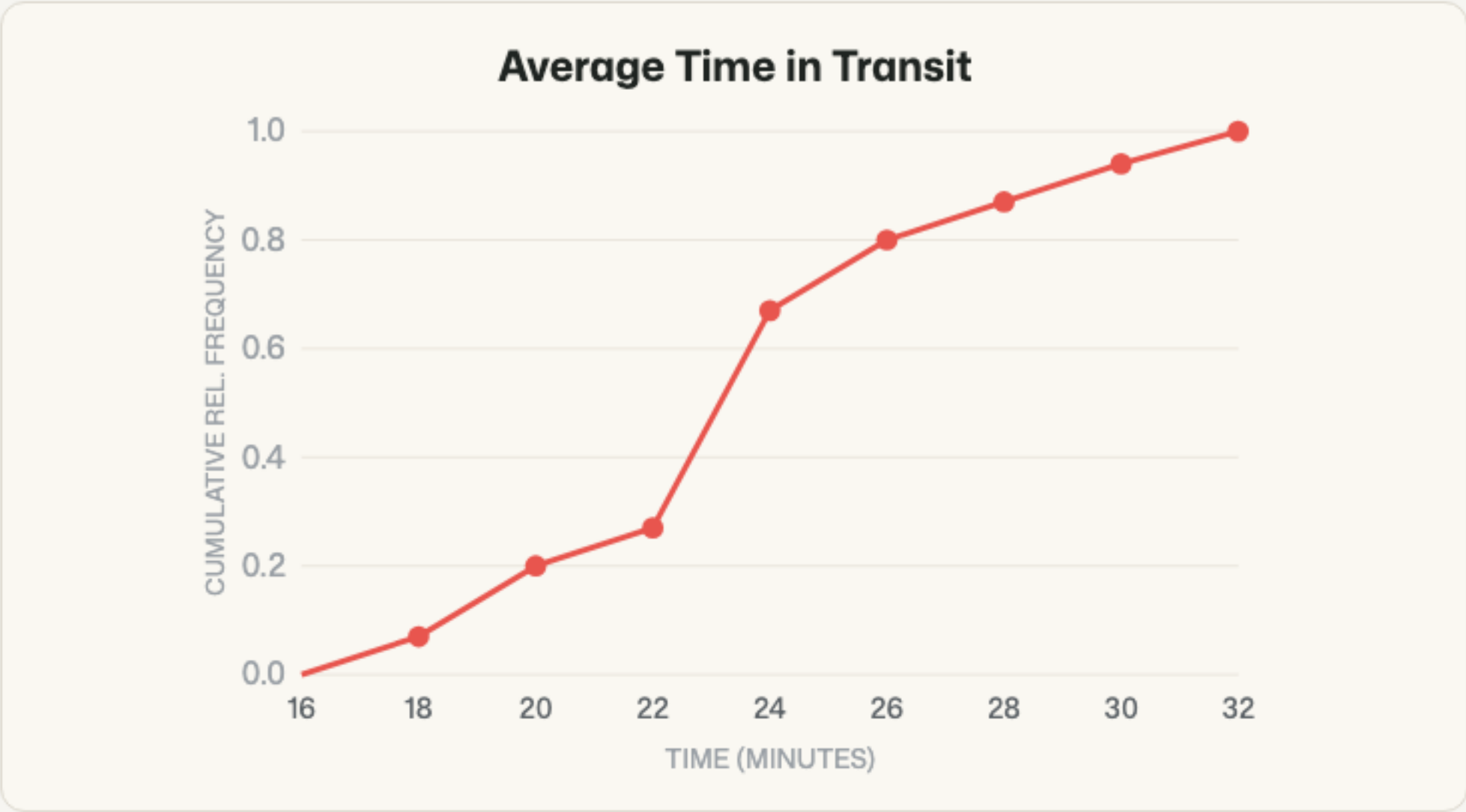
shows the **sum of the frequencies** up to and including the respective row.

Ogive

is the graph of the **cumulative relative frequency** across all classes.

⊖ Time in Transit · each row adds the previous relative frequency

⊖ TIME	RELATIVE FREQUENCY	CUMULATIVE REL. FREQUENCY
16-17.9	1/15 ≈ 0.07	0.07
18-19.9	2/15 ≈ 0.13	0.07+0.13 0.20
20-21.9	1/15 ≈ 0.07	0.20+0.07 0.27
22-23.9	6/15 ≈ 0.40	0.27+0.40 0.67
24-25.9	2/15 ≈ 0.13	0.67+0.13 0.80
26-27.9	1/15 ≈ 0.07	0.80+0.07 0.87
28-29.9	1/15 ≈ 0.07	0.87+0.07 0.94
30-31.9	1/15 ≈ 0.07	0.94+0.07 1.00



5

Data Analysis

5.1 Descriptive Statistics: Organizing and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

Measures of Central Tendency

Mean

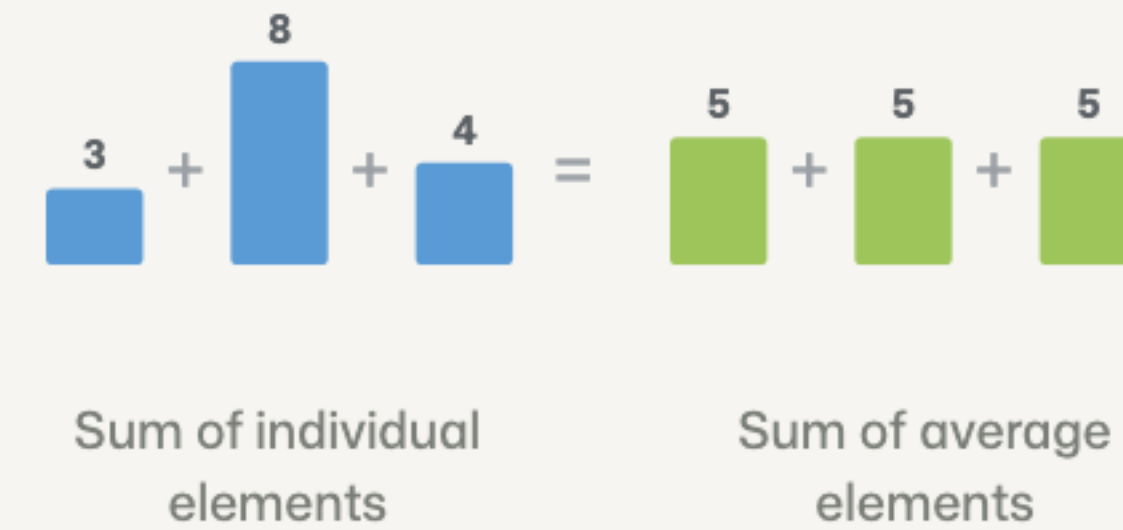
$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

ADVANTAGES

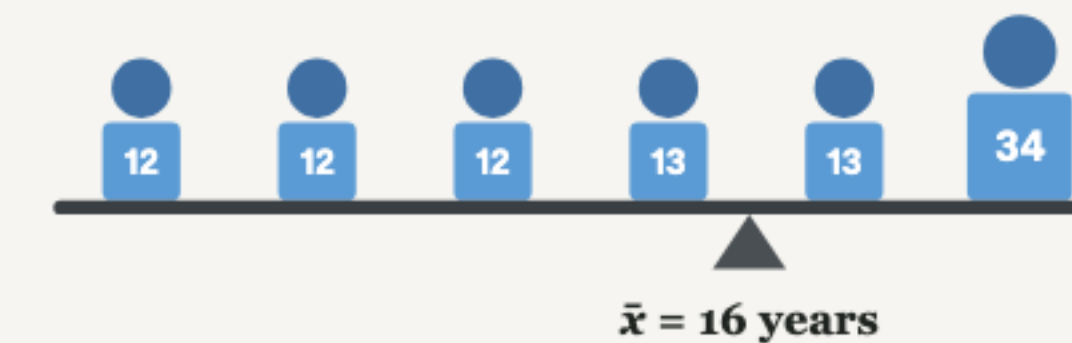
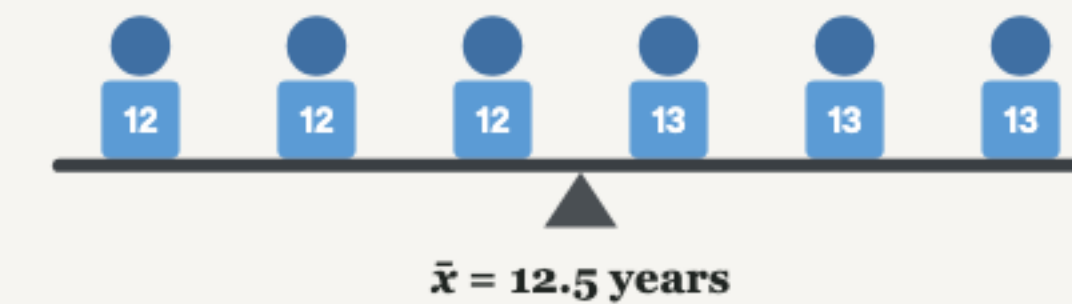
- Easy to compute: just sum up and divide.
- Intuitive – a single number “in the middle”; pulled up by large values and down by small ones.

DISADVANTAGES

- Can be **skewed by outliers** – poor for highly variable data.
- The mean of 100, 200 and –300 is 0. That is confusing.



The mean is the “center of gravity” – just like the balancing point



Measures of Central Tendency

Median

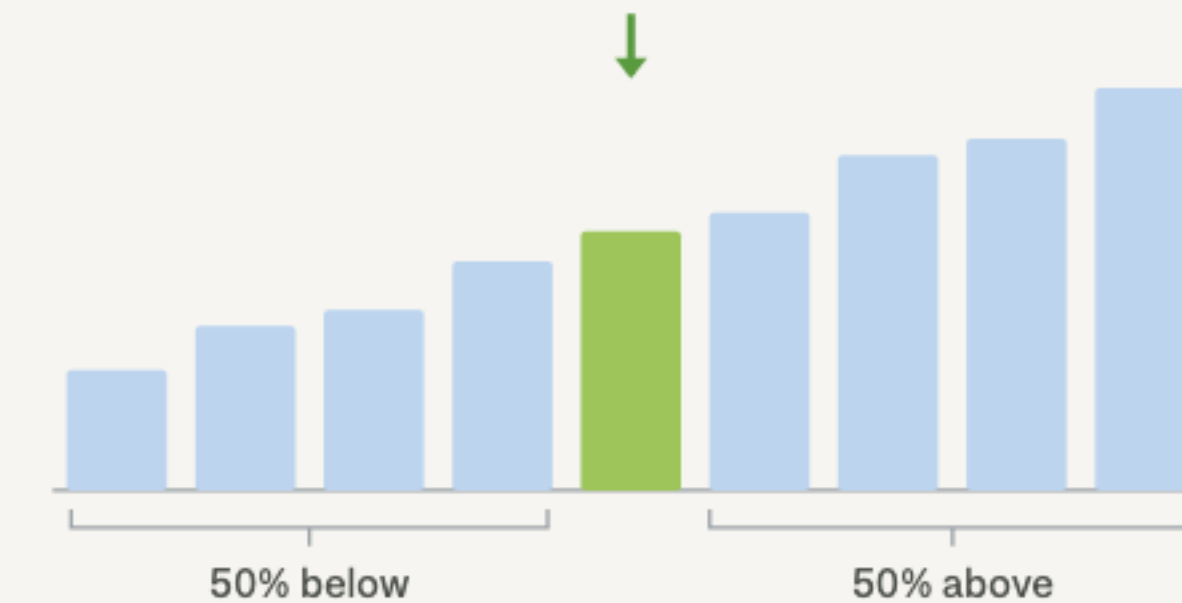
$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{for odd } n \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & \text{for even } n \end{cases}$$

ADVANTAGES

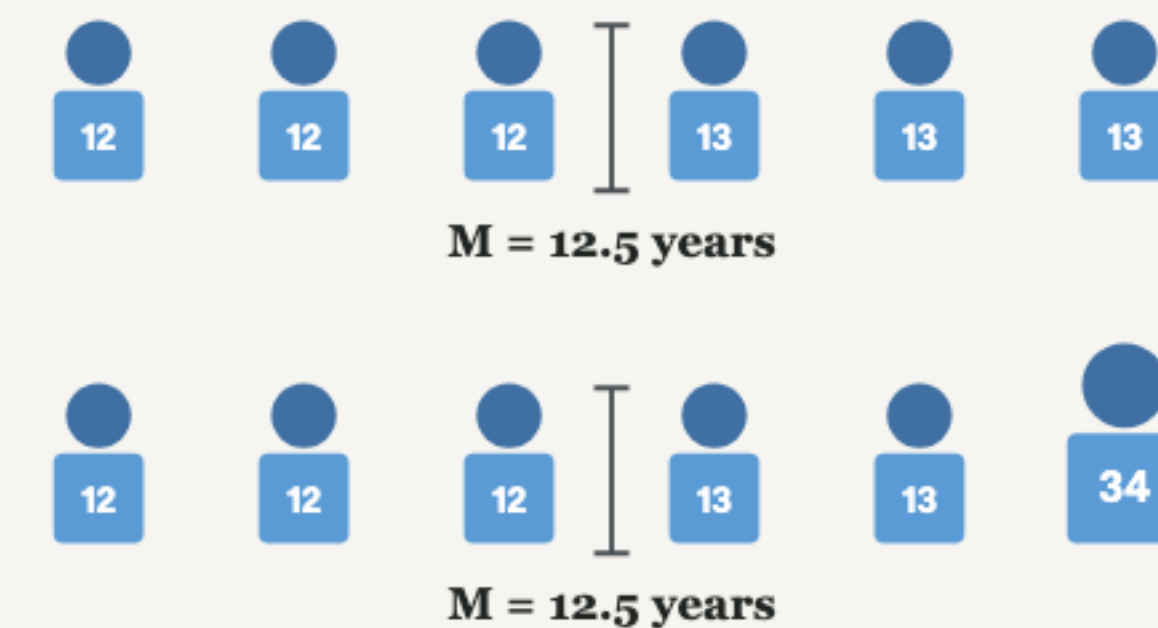
- Handles **outliers** well – often the most accurate depiction of a group.
- Splits the data into two equally sized groups.

DISADVANTAGES

- Harder to compute: data must first be **sorted**.
- Less well known; many confuse “median” with “average”.



The median is the middle element of a sorted list



Measures of Central Tendency

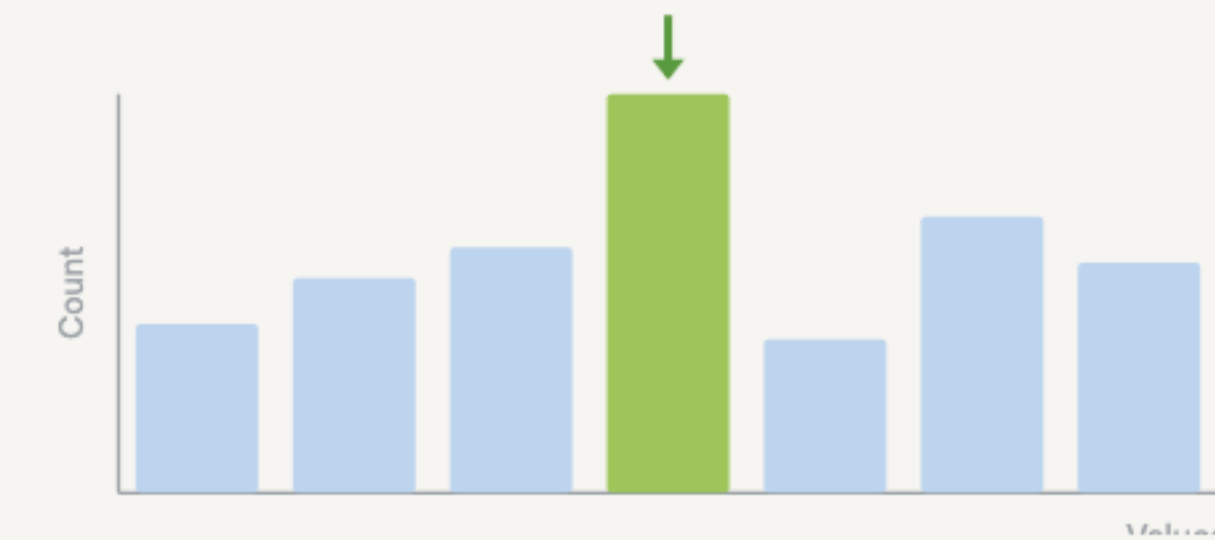
Mode

ADVANTAGES

- Good for exclusive choices (this one or the other; no compromises) – works with **nominal data**.
- Shows the choice that most people wanted (the mean often leads to a choice that no one wanted).
- Easy to understand.

DISADVANTAGES

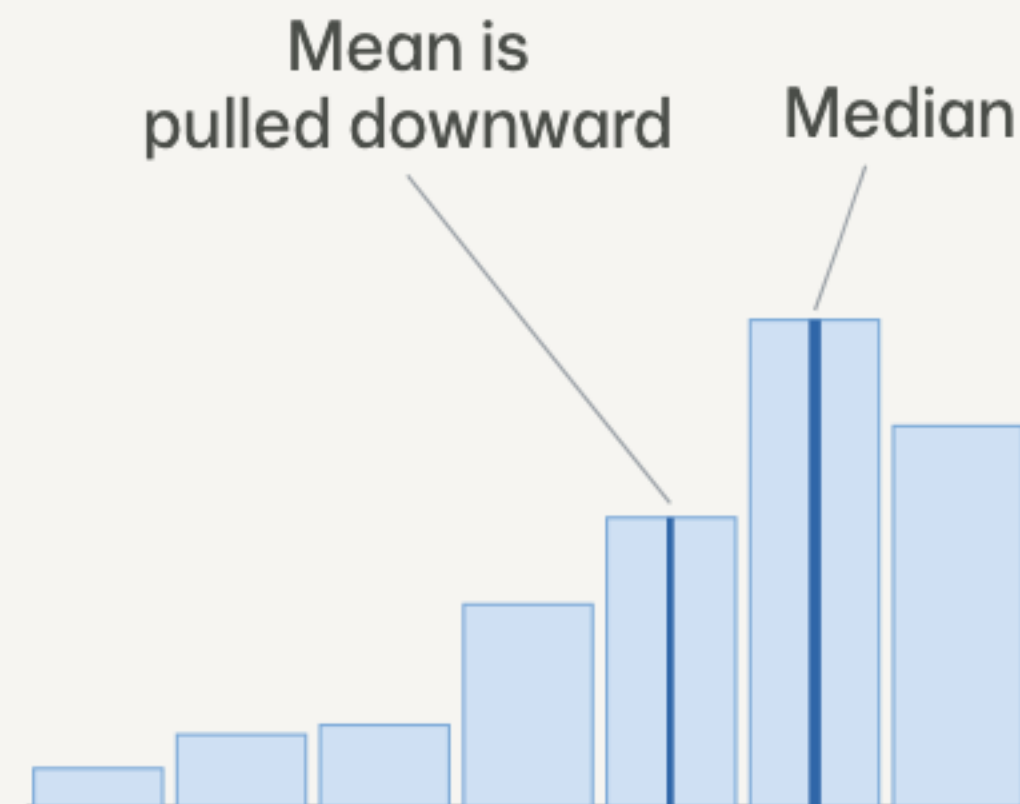
- Requires more effort: you have to count the votes.
- “The winner takes all” – there is no middle ground.



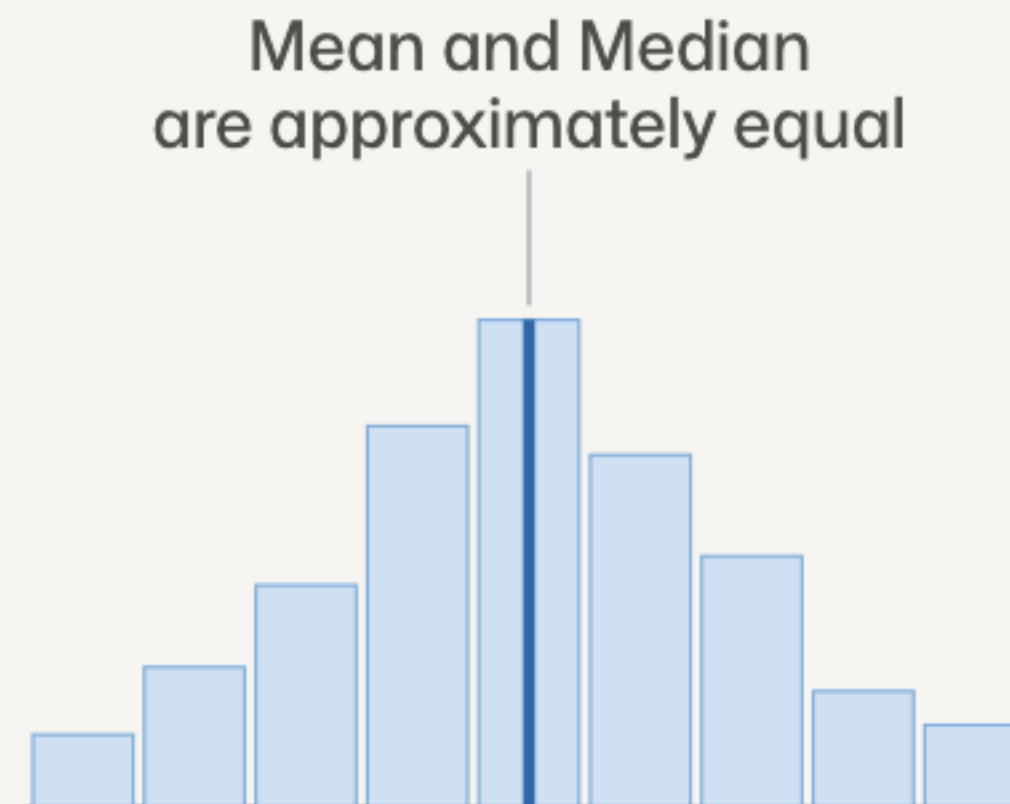
The mode is the most frequent value among all observations of the variable

The mode of       is 

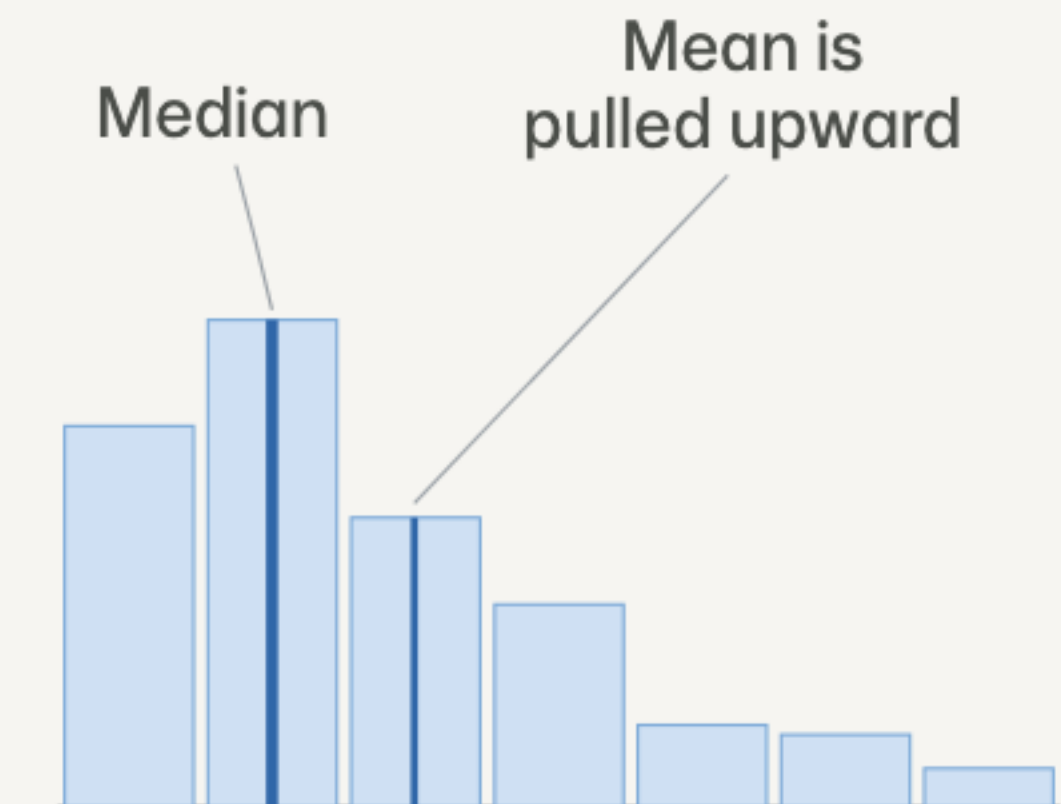
Measures of Central Tendency: Using Mean and Median to Identify the Distribution Shape



left-skewed



symmetric



right-skewed

Measures of Dispersion

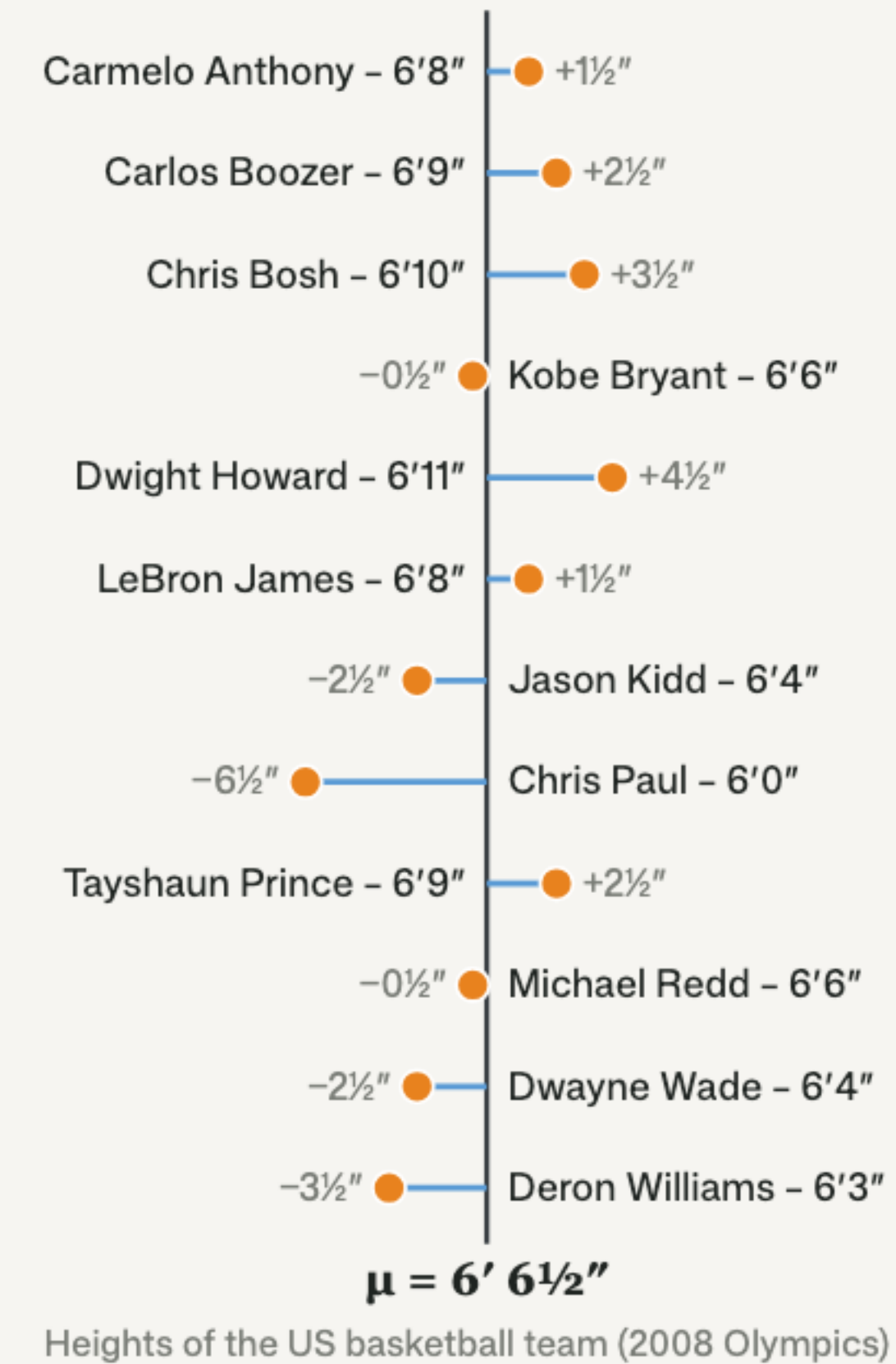
Variance is the average of squared deviations from the mean

Population Variance
(variance of the population)

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

Sample Variance

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$



Measures of Dispersion

Sample
Variance

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

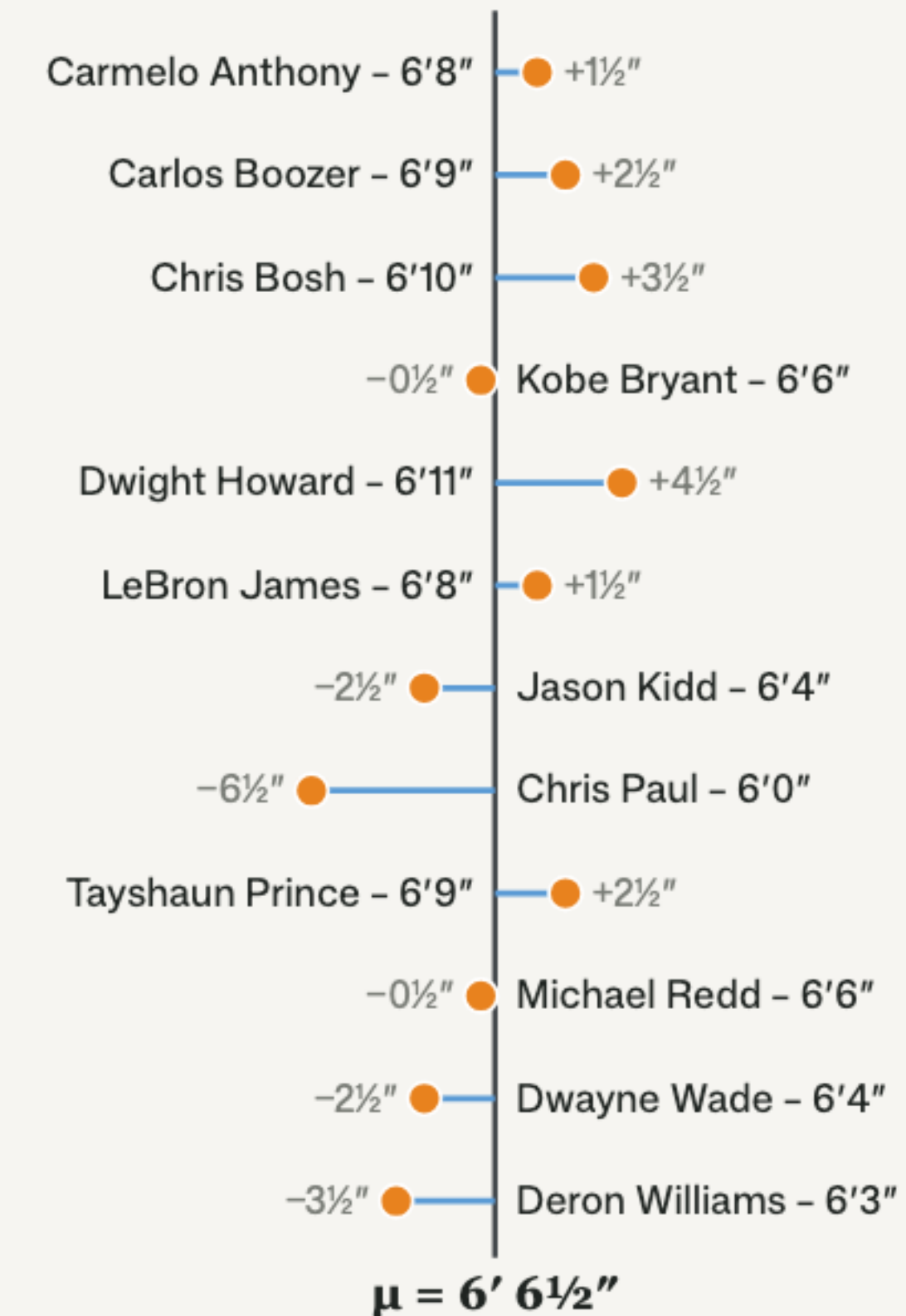
Why variance?

$$\bar{x} = \frac{1,5 + 2,5 + 3,5 - 0,5 + 4,5 + 1,5 - 2,5 - 6,5 + 2,5 - 0,5 - 2,5 - 3,5}{12} = 0$$

The mean acts like a **balance point** – the average deviation from the mean is always zero.

In the variance, all deviations are **squared**, so that negative and positive deviations do **not cancel out**.

$$s^2 = \frac{117}{12 - 1} \approx 10,6$$



Heights of the US basketball team (2008 Olympics)

Measures of Dispersion

Standard
Deviation

$$s = \sqrt{s^2}$$
$$\sigma = \sqrt{\sigma^2}$$

Standard deviation keeps the units of measurement of the original data.

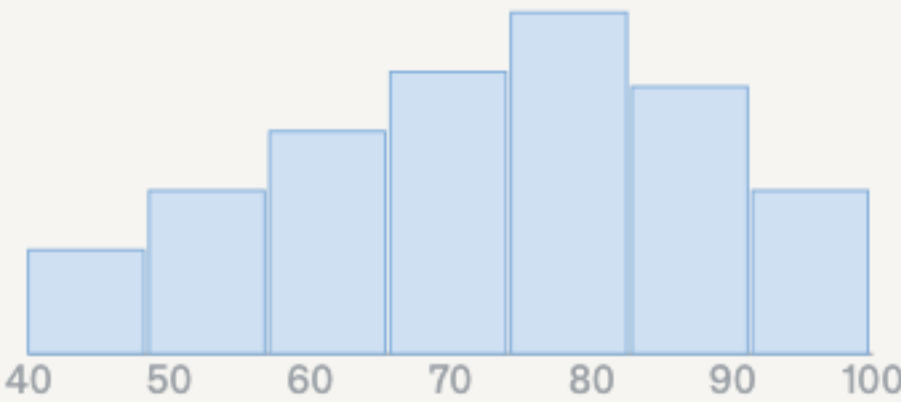
$$s^2 = \frac{117}{12 - 1} \approx 10,6$$

square inches

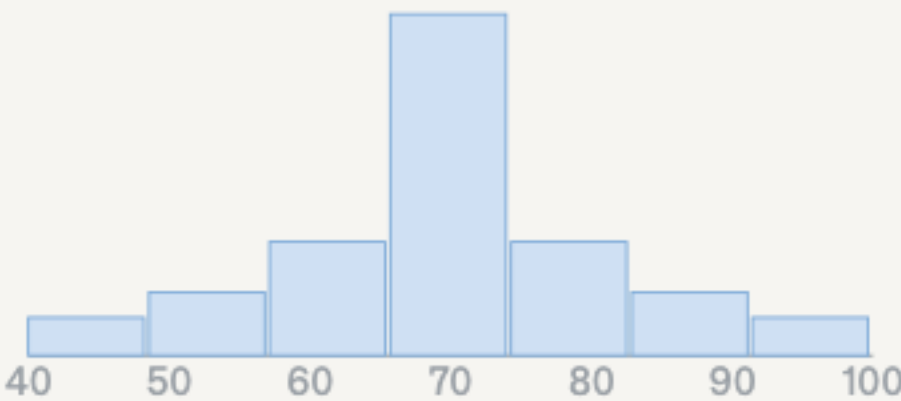
$$s = \sqrt{10,6} \approx 3,3$$

inches

Which data set has a higher standard deviation?

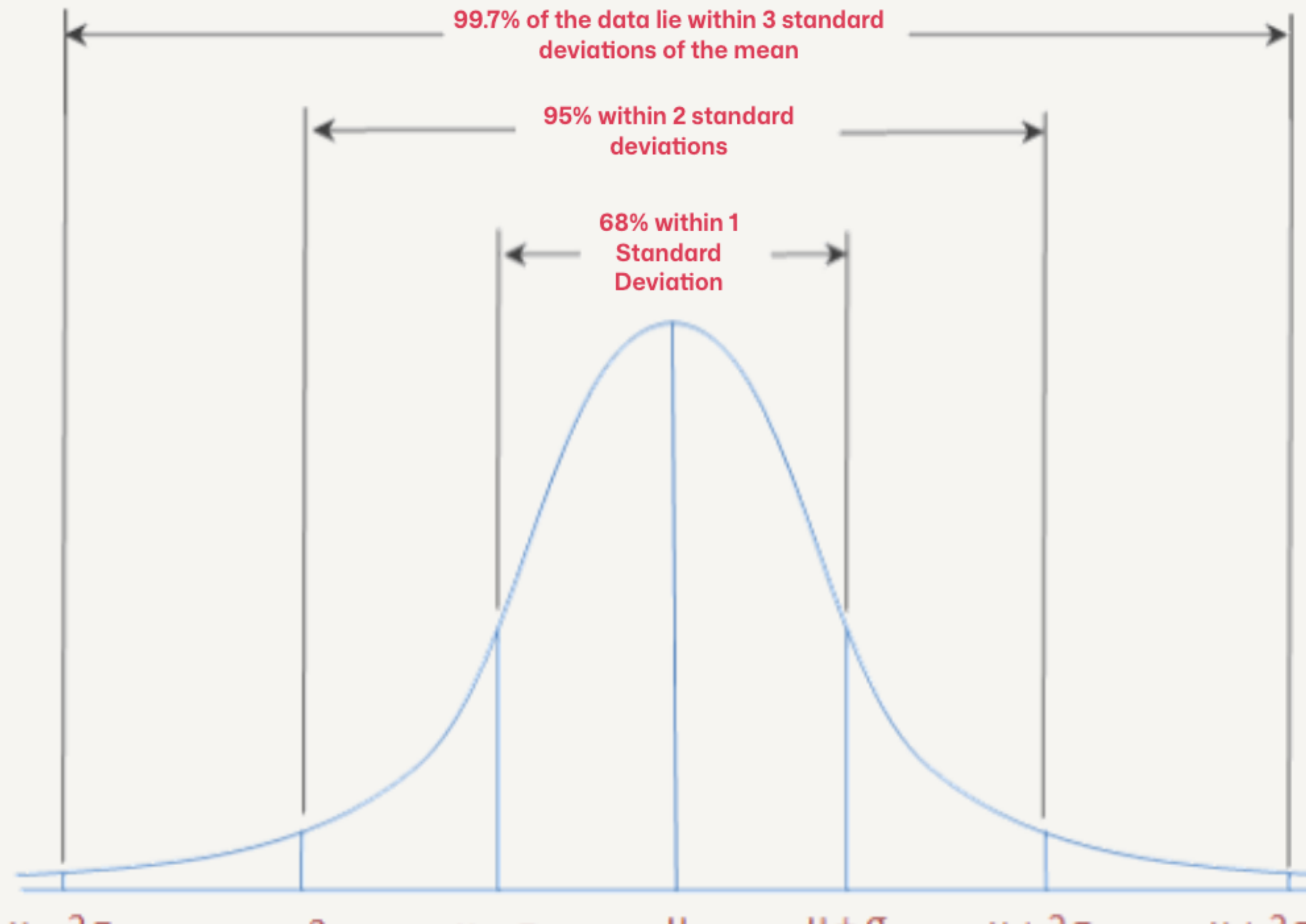


48 49 52 55 57 58
62 64 65 66 67 72
72 73 75 78 78 78
79 82 84 86 88 89
93 94 95



48 55 57 61 64 65
68 71 71 72 73 73
74 75 78 78 79 79
79 79 82 84 85 88
89 92 95

Relationship between the Standard Deviation and the Shape of the Normal Distribution



5

Data Analysis

5.1 Descriptive Statistics: Organizing and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

Cross-Tabulations



Cross-Tabulations

Cross-tabulations summarize the joint distribution of two (or more) discrete variables in a table.

They help analyze the **relationship** of one variable (e.g. brand loyalty) with another (e.g. gender).

Each cell represents a **combination of the categories**.

Typical questions a cross-tabulation answers:

- How many brand-loyal consumers are men?
- Is the usage frequency (high, medium, low) of a product related to outdoor activities (often, sometimes, rarely, never)?
- Is familiarity with a new product related to age and level of education?
- Is ownership of a product related to income (high, medium, low)?

Cross-Tabulations

Does ownership of expensive car brands depend on level of education?

Ownership of expensive car brands by level of education

Ownership of an expensive car	Level of education	
	College degree	No college degree
yes	32 %	21 %
no	68 %	79 %
Total	100 %	100 %
Number of cases	250	750

Cross-Tabulations

Sometimes introducing a **third variable** can reveal ...

CASE 1

Spurious relationships

The apparent relationship disappears once you control for it.

CASE 2

Hidden relationships

A previously invisible relationship becomes visible.

CASE 3

No change

The original relationship remains unchanged.

Cross-Tabulations

CASE 1
Spurious
relationship

Does ownership of expensive car brands depend on level of education?

Ownership of expensive car brands by level of education and income level

Ownership of an expensive car	High income		Low income	
	College degree	No college degree	College degree	No college degree
yes	20 %	20 %	40 %	40 %
no	80 %	80 %	60 %	60 %
Total	100 %	100 %	100 %	100 %
Number of cases	100	700	150	50

Is the relationship still there?

No – within each income group, car ownership is the same (20% and 40% respectively). The relationship with level of education was **spurious**; in fact, income is what matters.

Cross-Tabulations

CASE 2
Hidden
Relationship

Does age influence the desire to travel and seek adventure?

Desire for international travel by age

Desire for international travel	Age	
	Under 45	45 and over
yes	50 %	50 %
no	50 %	50 %
Total	100 %	100 %
Number of cases	500	500

... by age and gender

Desire for international travel	Male		Female	
	< 45	≥ 45	< 45	≥ 45
yes	60 %	40 %	35 %	65 %
no	40 %	60 %	65 %	35 %
Total	100 %	100 %	100 %	100 %
Number of cases	300	300	200	200

Aggregated (50%/50%), age appears to play no role. Only when split by gender does it become clear: for men the desire to travel rises with age, for women it falls – a **hidden relationship**.

Cross-Tabulations

CASE 3
No
Change

Is the frequency of visits to fast-food restaurants related to family size?

Visit frequency by family size

Frequently go to fast-food restaurants	Family size	
	Small	Large
yes	50 %	50 %
no	50 %	50 %
Total	100 %	100 %
Number of cases	500	500

... by family size and income

Frequently go to fast-food restaurants	Low income		High income	
	Small	Large	Small	Large
yes	50 %	50 %	50 %	50 %
no	50 %	50 %	50 %	50 %
Total	100 %	100 %	100 %	100 %
Number of cases	250	250	250	250

Even after introducing the third variable (income), it stays 50%/50% everywhere – the original (non-)relationship **does not change**.

5

Data Analysis

5.1 Descriptive Statistics: Displaying and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

5

Data Analysis

5.1 Descriptive Statistics: Displaying and Presenting Data

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

Hypothesis Testing

Hypothesis Testing

A **five-step procedure** that, based on a sample and using probability theory, determines whether a hypothesis is sufficiently supported.

In other words: a method for testing whether the results of a random sample can be **generalized to the population**.

A five-step approach:

- Formulating a **null hypothesis** and its **alternative hypothesis**
- Setting the **significance level**
- Choosing the appropriate **test statistic**
- Formulating the **decision rule**
- Calculating the metrics from the sample and **making the decision**

"People are mistakenly confident in their knowledge and underestimate the likelihood that their beliefs will turn out to be wrong. They tend to seek out only information that confirms what they already believe."

– Max Bazerman

Hypothesis Testing

Do men really use the internet more often than women – across the whole population?

Internet use and gender · Sample n = 30

Internet use	Gender		Total
	Male	Female	
rarely	5	10	15
frequently	10	5	15
Total	15	15	n = 30

In the sample, men use the internet more frequently. But does this also hold in the **population** – or is it a coincidence of the sample?

Hypothesis Testing

Step 1 · Formulating a null hypothesis and its alternative hypothesis

H_0 There is **no difference** between men and women in the frequency of internet use.

$$IN_m = IN_f$$

H_1 Men and women show **different** internet use behavior.

$$IN_m \neq IN_f$$

Null hypothesis (H_0) is a claim of the status quo – that there is no difference or no effect.

Alternative hypothesis (H_1) claims the opposite – that there is a difference or an effect.

Hypothesis Testing

Step 2 · Setting the significance level

Significance (α) – probability that a **true** null hypothesis is rejected.

β – probability that a **false** null hypothesis is accepted.

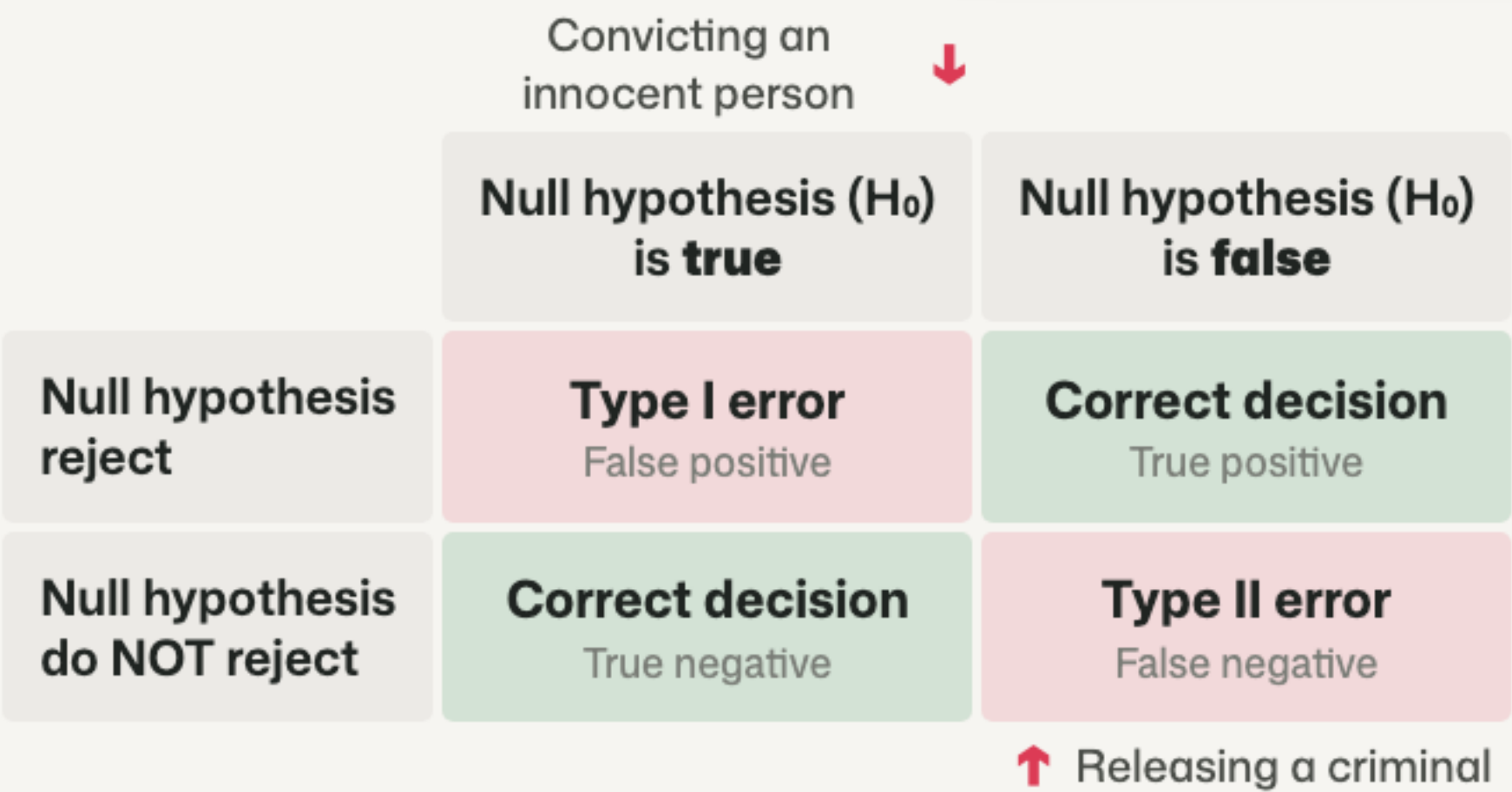
	α – Significance ↓	
	Null hypothesis (H_0) is true	Null hypothesis (H_0) is false
Null hypothesis reject	Type I error False positive	Correct decision True positive
Null hypothesis do NOT reject	Correct decision True negative	Type II error False negative
		↑ $(1 - \beta)$ – Power

Hypothesis Testing

Step 2 · Setting the significance level

Analogy: **innocence in a criminal trial.**
H₀: The defendant is innocent.

Significance (α) – probability that a **true** null hypothesis is rejected.
 β – probability that a **false** null hypothesis is accepted.



Hypothesis Testing

Step 2 · Setting the significance level

Analogy: **a rustling in the bushes – is that a lion?**

H_0 : There is no lion in the bushes.

Significance (α) – probability that a **true** null hypothesis is rejected.

β – probability that a **false** null hypothesis is accepted.

	There is no lion, but you run away ↓	
	Null hypothesis (H_0) is true	Null hypothesis (H_0) is false
Null hypothesis reject	Type I error False positive	Correct decision True positive
Null hypothesis do NOT reject	Correct decision True negative	Type II error False negative
		↑ You stay unconcerned – and the lion eats you

Hypothesis Testing

Step 2 · Setting the significance level

Significance (α) – the probability that a **true** null hypothesis is rejected.

β – the probability that a **false** null hypothesis is accepted.

Significance levels in market research

α – significance level

$(1 - \alpha)$ – confidence level

0.01 (1%)

0.99 (99%)

0.05 (5%)

0.95 (95%)

Hypothesis Testing

Step 3 · Choosing the appropriate test statistic

Our example is about the **distribution of non-metric variables** (rare/frequent internet use; men/women) in **one sample**.

SAMPLE	APPLIED TO	SCALE LEVEL	TEST STATISTICS / COMMENTS
One sample	Distributions	Non-metric	Kolmogorov-Smirnov and χ^2 test for goodness of fit; runs test for randomness; binomial test for dichotomous variables
	Means	Metric	t-test (variance unknown); z-test (variance known)
	Proportions	Metric	z-test
Two independent samples	Distributions	Non-metric	Kolmogorov-Smirnov test for agreement of distributions between two samples
	Means	Metric	Two-sample t-test; F-test for equality of variances
	Proportions	Metric, Non-metric	z-test; χ^2 test
	Ranks / Medians	Non-metric	Mann-Whitney U-test (more sensitive than the median test)
Paired samples	Means	Metric	Paired-difference t-test
	Proportions	Non-metric	McNemar test for binary variables; χ^2 test
	Ranks / Medians	Non-metric	Wilcoxon signed-rank test (more sensitive than the sign test)

Hypothesis Testing

Step 3 · Choosing the appropriate test statistic

One sample · distribution · non-metric → the **χ^2 test for goodness of fit**.

SAMPLE	APPLIED TO	SCALE LEVEL	TEST STATISTICS / COMMENTS
One sample	Distributions	Non-metric	Kolmogorov-Smirnov and χ^2 test for goodness of fit ; runs test for randomness; binomial test for dichotomous variables
	Means	Metric	t-test (variance unknown); z-test (variance known)
	Proportions	Metric	z-test
Two independent samples	Distributions	Non-metric	Kolmogorov-Smirnov test for agreement of distributions between two samples
	Means	Metric	Two-sample t-test; F-test for equality of variances
	Proportions	Metric, Non-metric	z-test; χ^2 test
	Ranks / Medians	Non-metric	Mann-Whitney U-test (more sensitive than the median test)
Paired samples	Means	Metric	Paired-difference t-test
	Proportions	Non-metric	McNemar test for binary variables; χ^2 test
	Ranks / Medians	Non-metric	Wilcoxon signed-rank test (more sensitive than the sign test)

Hypothesis Testing

Step 3 · Choosing the appropriate test statistic

The χ^2 test statistic (**chi-square**) tests the statistical significance of the relationship observed in a crosstab.

H_0 There is **no relationship** between the variables.

χ^2 tests the **equality of frequency distributions** – which frequencies are we comparing?

- f_e – frequencies we **would expect** in the cells if there were no relationship.
- f_o – the frequencies actually **observed**.

Hypothesis Testing

Step 3 · Choosing the appropriate test statistic

f_e – frequencies we would expect in the cells if there were no relationship.

f_o – the frequencies actually observed.

$$f_e = \frac{n_r \cdot n_c}{n}$$

n_r – total sum in a **row**

n_c – total sum in a **column**

n – sample size

$$f_{e1,1} = \frac{15 \cdot 15}{30} = 7.5$$

$$f_{e2,1} = \frac{15 \cdot 15}{30} = 7.5$$

$$f_{e1,2} = \frac{15 \cdot 15}{30} = 7.5$$

$$f_{e2,2} = \frac{15 \cdot 15}{30} = 7.5$$

Hypothesis Testing

f_e – frequencies we would expect in the cells if there were no relationship.

f_o – the frequencies actually observed.

Step 3 · Choosing the appropriate test statistic

$$\chi^2 = \sum_{\text{Alle Zellen}} \frac{(f_o - f_e)^2}{f_e}$$

In our example:

$$\chi^2 = \frac{(5-7.5)^2}{7.5} + \frac{(10-7.5)^2}{7.5} + \frac{(10-7.5)^2}{7.5} + \frac{(5-7.5)^2}{7.5} = 0.833 \cdot 4 = \mathbf{3.333}$$

χ^2 should always be computed using only **absolute frequencies**. If the data are given in percentages (relative frequencies), they must first be converted into absolute frequencies.

Hypothesis Testing

Step 4 · Formulating the decision rule

TS_{cal} – the observed (calculated) value of the test statistic.

TS_{cr} – the critical value of the test statistic for the chosen significance level.

If the probability of **TS_{cal}** < significance level (α), then reject **H₀**.

or

If **TS_{cal}** > **TS_{cr}**, then reject **H₀**.

Hypothesis Testing

Step 4 · Formulating the decision rule

Critical values of χ^2 for various significance levels α

df	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	–	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209

$df = (r-1)(c-1)$
df – degrees of freedom
r – number of rows
c – number of columns
 $df = (2-1)(2-1) = 1$

If the probability of TS $_{cal} < (\alpha)$, then reject H_o .

or

If TS $_{cal} > TS_{cr}$, then reject H_o .

$\chi^2_{cal} = 3.333$

$\chi^2_{cr} = 3.841$

$3.333 < 3.841 \Rightarrow \chi^2_{cal} < \chi^2_{cr}$

H_o CANNOT be rejected

Hypothesis Testing

Step 5 · Making the decision

Is the evidence there?

What are the consequences?

- H_0 – that there is no relationship – **cannot be rejected**.
- The relationship is **0.05** statistically **not significant**.
- The results observed in the sample **cannot** be generalized to the population.

Hypothesis Testing

Internet use and gender · $n = 30$

Internet use	Gender		Total
	Male	Female	
rarely	5	10	15
frequently	10	5	15
Total	15	15	$n = 30$

Do men really use the internet more often than women – in the population?

Answer: The sample provides no evidence for it.

If the sample was carefully selected and drawn, we can claim with **95% confidence** that there is no such relationship.

Otherwise – **we don't know.**

5

Data Analysis

5.1 Descriptive Statistics: A Concise Overview of Statistical Techniques

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

5.2.3 Relationship between Two (Ratio Scaled) Variables

Testing the Strength of a Relationship

χ^2 tests only the **significance** of a relationship and says nothing about its **strength**.

Simple proof: doubling all the values in the cross-tabulation doubles χ^2 – but the relationship itself stays the same.

Measures of the strength of a relationship are:

 φ

Phi Coefficient

 C

Contingency Coefficient

 V

Cramer's V

 λ

Lambda Coefficient

Phi Coefficient

$$\varphi = \sqrt{\frac{\chi^2}{n}}$$

The higher φ , the stronger the relationship between the variables.

Values **> 0.30** are considered substantial.

Problems:

- φ is not standardized and has an upper limit of 1 only for 2×2 tables; it depends on the table dimensions.
- φ values from different studies cannot be compared with one another.

In our example:

$$\varphi = \sqrt{\frac{3,333}{30}} = 0,333$$

The relationship is **not particularly strong**.

Contingency Coefficient

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

The higher C , the stronger the relationship between the variables.

Values **> 0.30** are considered substantial.

Although C values have an upper limit of 1, they cannot actually reach this limit.

Problems:

- C is not standardized and depends on the table dimensions.
- C values from different studies cannot be compared with one another.

In our example:

$$C = \sqrt{\frac{3,333}{3,333 + 30}} = 0,316$$

The relationship is **not particularly strong**.

Cramer's V

$$V = \sqrt{\frac{\chi^2}{n \cdot (\min(r, c) - 1)}}$$

r – number of rows

c – number of columns

The higher V , the stronger the relationship between the variables.

Values **> 0.30** are considered substantial.

V values have an upper limit of 1, but they too can actually reach it only for 2×2 tables.

Problems:

- V is not standardized and depends on the table dimensions.
- V values from different studies cannot be compared with one another.

In our example:

$$V = \sqrt{\frac{3.333}{30 \cdot (2 - 1)}} = 0.333$$

The relationship is **not particularly strong**.

Lambda Coefficient

$$\lambda = \frac{\sum_c \max_r(n_{rc}) - \max_r(n_r)}{n - \max_r(n_r)}$$

r – row index

c – column index

Indicates the extent to which knowing the value of one variable helps in **predicting** the other variable.

Is standardized between **0 and 1** (1 – error-free prediction, 0 – no improvement in prediction).

λ values from different studies **can** be compared with one another.

In our example:

$$\lambda = \frac{(10 + 10) - 15}{30 - 15} = 0.333$$

Knowing the gender increases prediction accuracy by a factor of **0.333**, i.e. **33.3% improvement**.

Lambda Coefficient

Sum of the maximum frequencies
of all columns

$$\lambda = \frac{\sum_c \max_r(n_{rc}) - \max_r(n_r)}{n - \max_r(n_r)}$$

maximum total value of a row

r – row index

c – column index

	Internet Use	Gender		Total (row)
		Male	Female	
$r = 1$	rarely	5	10	15
$r = 2$	often	10	5	15
	Total (column)	15	15	$n = 30$
		$c = 1$	$c = 2$	

$$\lambda = \frac{(10 + 10) - 15}{30 - 15} = 0.333$$

Knowing the gender increases prediction accuracy by a factor of **0.333**,
i.e. **33.3% improvement**.

5

Data Analysis

5.1 Descriptive Statistics: A Concise Overview of Statistical Techniques

5.1.1 Organizing Qualitative Data

5.1.2 Organizing Quantitative Data

5.1.3 Summarizing Data Numerically

5.1.4 Cross-Tabulations

5.2 Inferential Statistics: Generalizing to the Population

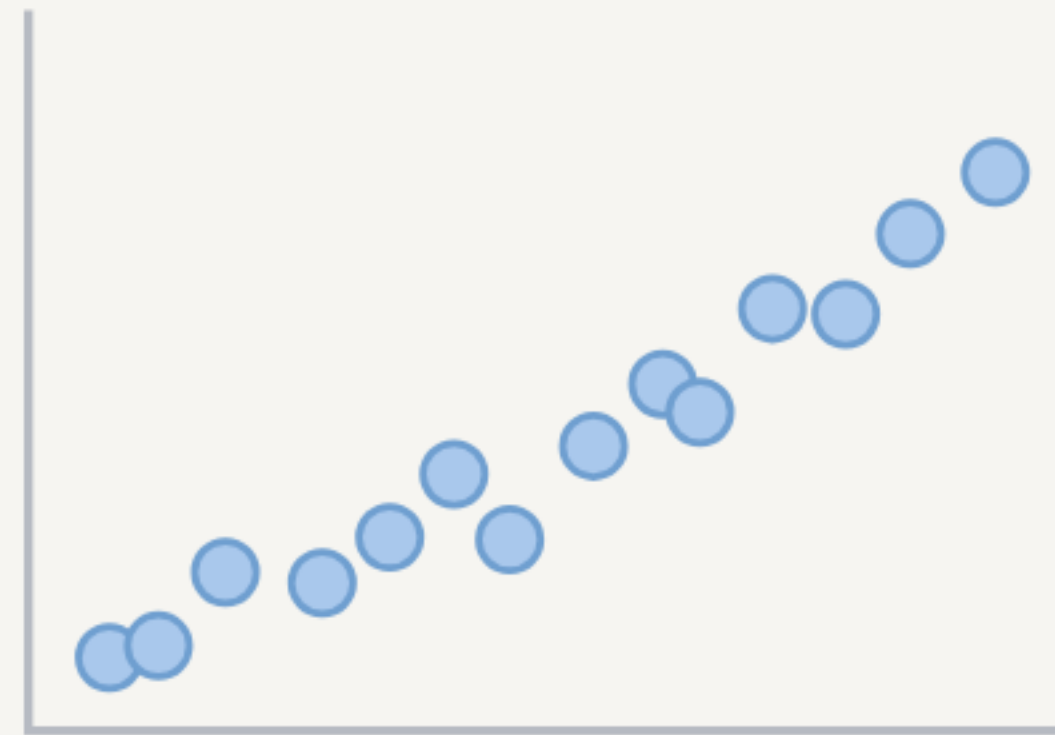
5.2.1 Hypothesis Testing

5.2.2 Strength of a Relationship in Cross-Tabulation

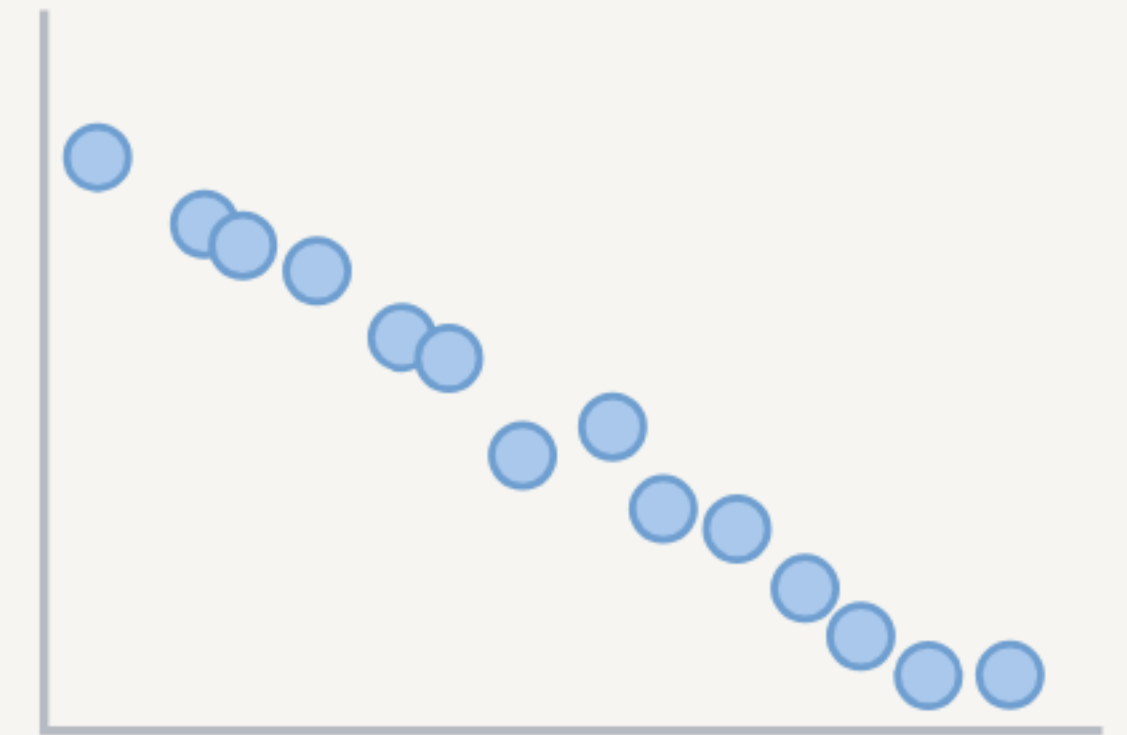
5.2.3 Relationship between Two (Ratio Scaled) Variables

Types of Relationships between Two Variables

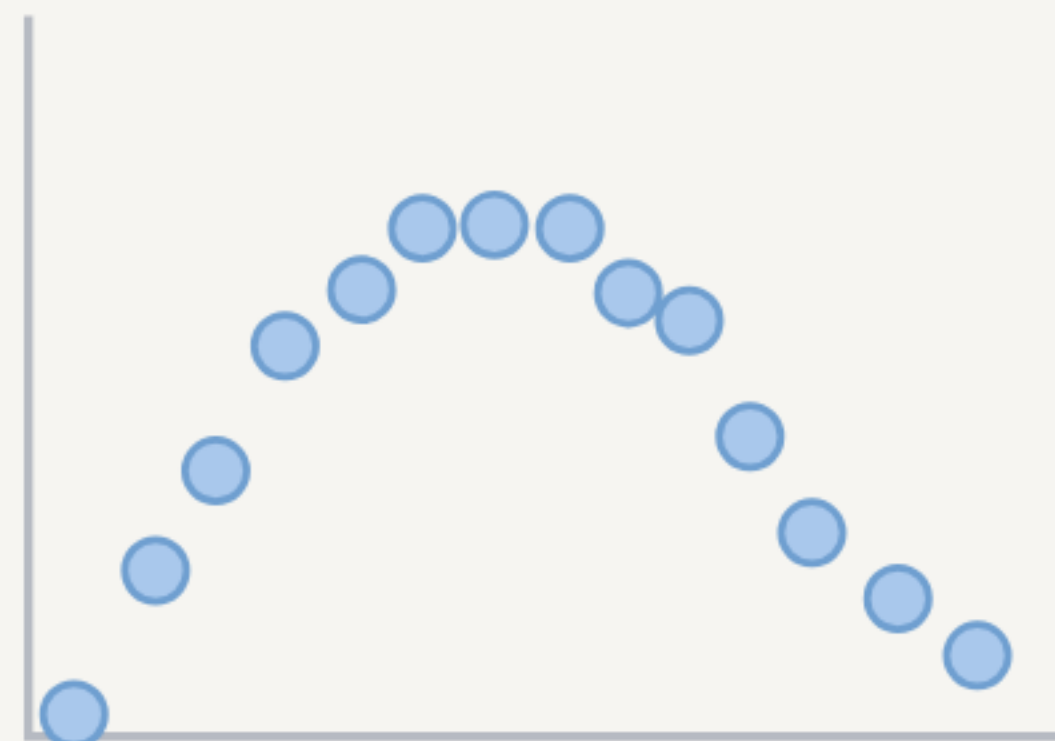
Unless the data come from a controlled experiment, we can only claim the **existence** of a relationship between the variables – but not the **causal direction** of that relationship.



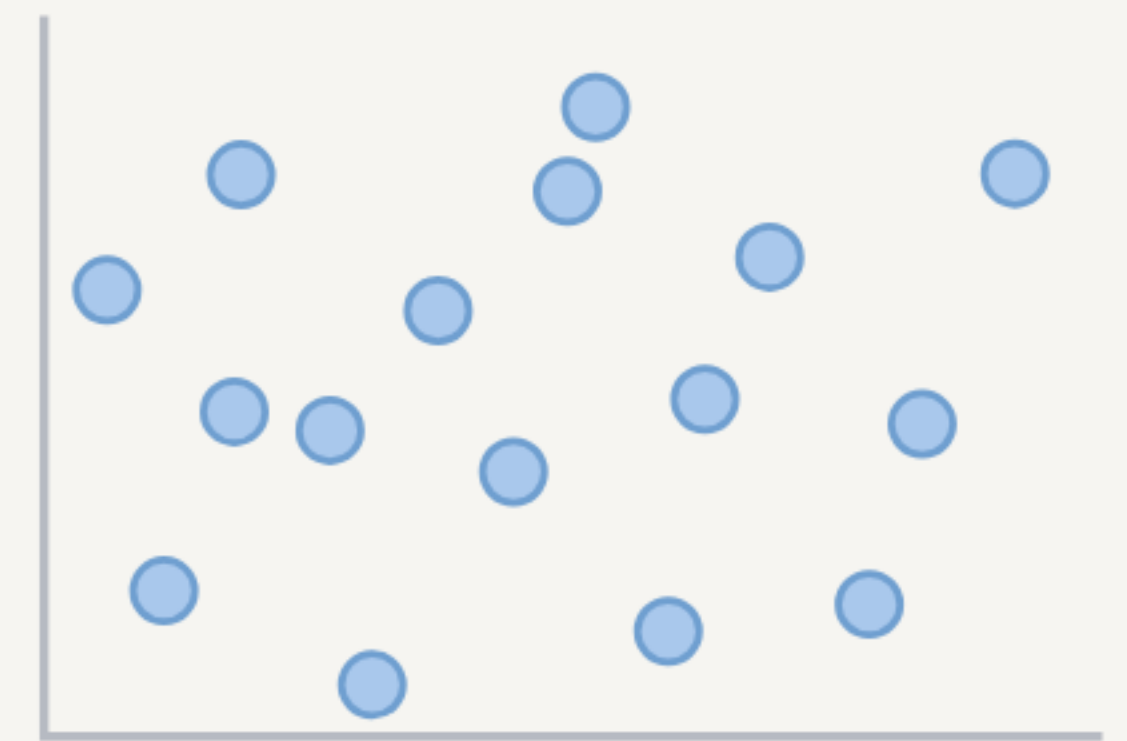
Linear



Linear



Non-linear



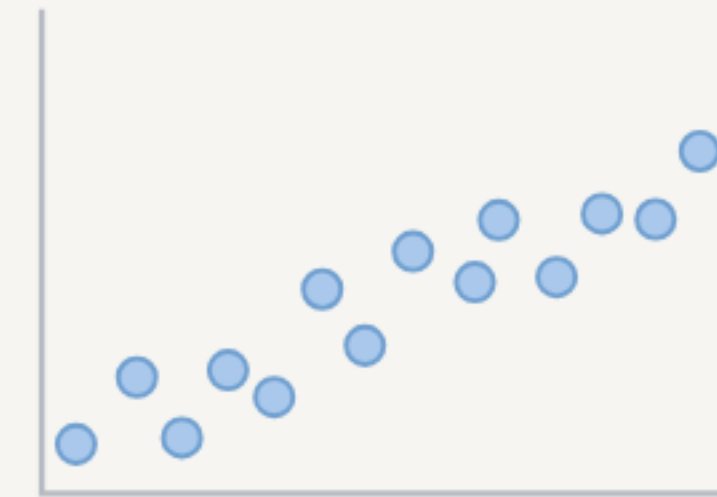
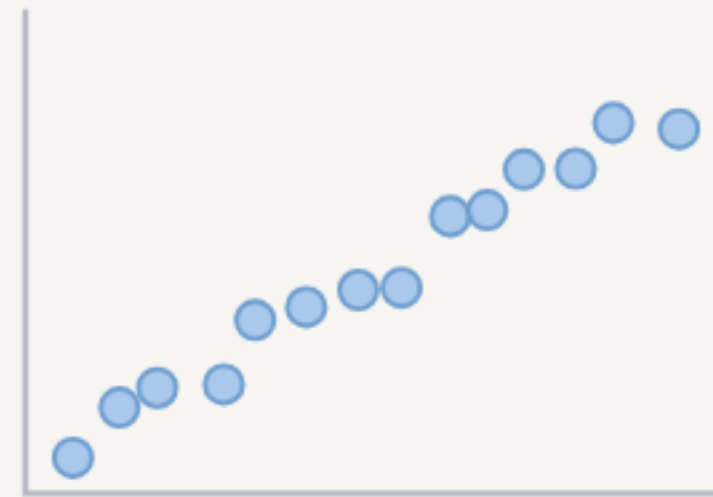
No relationship

Linear Correlation

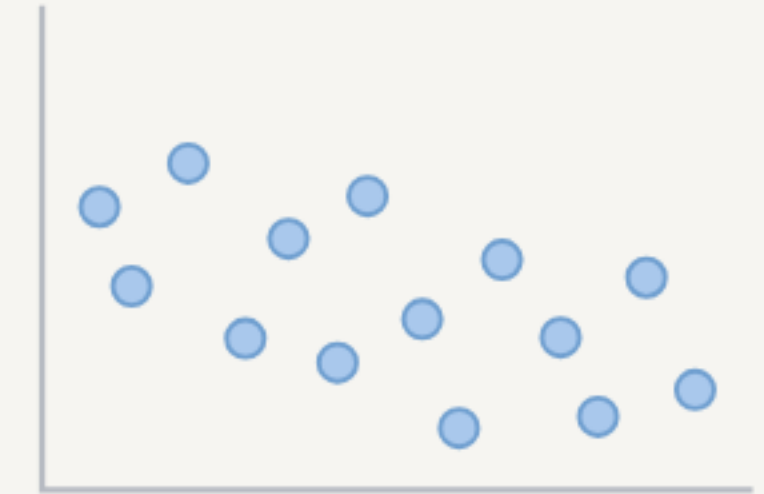
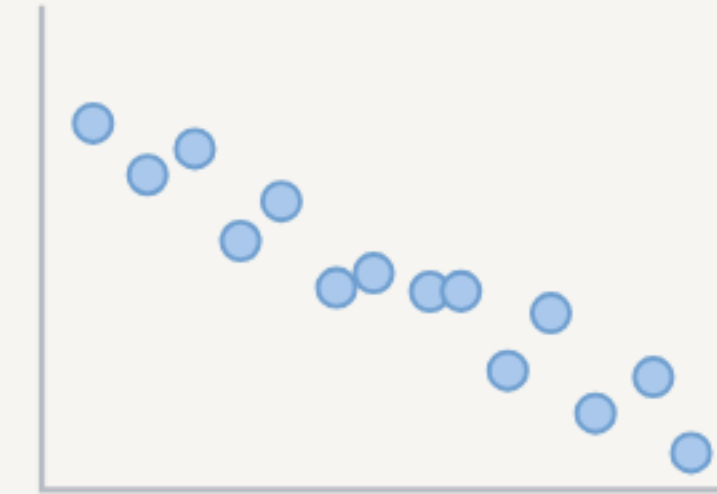
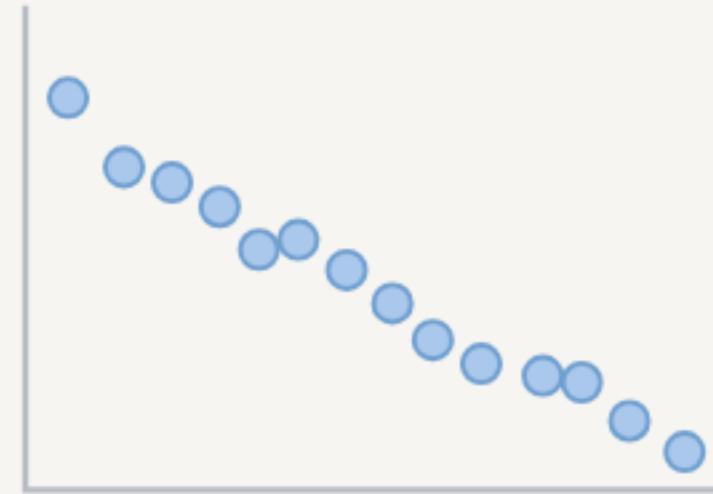
Two variables **correlate positively** when higher values of one variable correspond to higher values of the other variable.

Two variables **correlate negatively** when higher values of one variable correspond to lower values of the other variable.

Positive Correlation



Negative Correlation



Linear Correlation Coefficient

Properties:

- Values of the linear correlation coefficient always lie between **-1 and 1**.
- When $r = +1$ there is a perfect **positive** linear relationship between the variables.
- When $r = -1$ there is a perfect **negative** linear relationship between the variables.
- The closer r is to +1 or -1, the stronger the respective relationship.
- If r is close to **0**, there is little evidence of a linear relationship – but this does not mean there is no relationship at all, just no linear.

Linear Correlation Coefficient

The **(Pearson) linear correlation coefficient** measures the strength of the linear relationship between two variables.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$$

Linear Correlation Coefficient

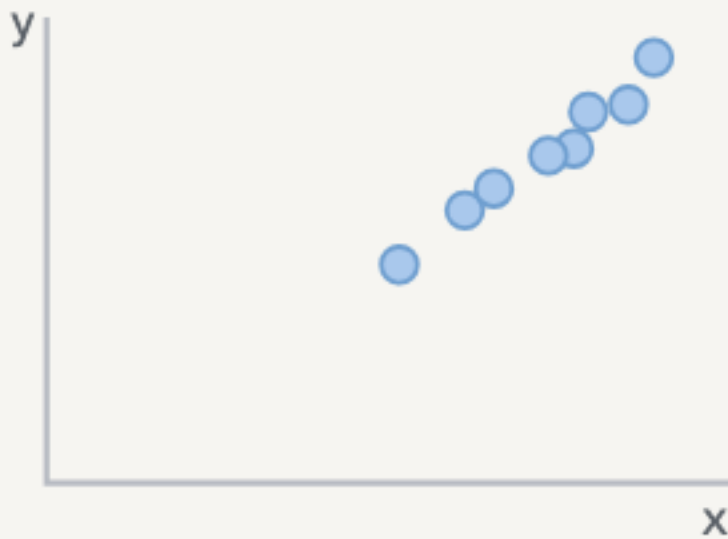
Strength of the Relationship between Variables

r-value	Interpretation
0 to 0.3	Very weak
0.3 to 0.5	Weak
0.5 to 0.7	Moderate
0.7 to 0.9	High
0.9 to 1	Very high

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$$

x	y	(x _i - $\bar{x}$)	(y _i - $\bar{y}$)	(x _i - $\bar{x}$)(y _i - $\bar{y}$)	(x _i - $\bar{x}$) ²	(y _i - $\bar{y}$) ²
86	98	12.5	13.5	168.75	156.25	182.25
62	70	-11.5	-14.5	166.75	132.25	210.25
52	56	-21.5	-28.5	612.75	462.25	812.25
90	110	16.5	25.5	420.75	272.25	650.25
66	76	-7.5	-8.5	63.75	56.25	72.25
80	96	6.5	11.5	74.75	42.25	132.25
78	86	4.5	1.5	6.75	20.25	2.25
74	84	0.5	-0.5	-0.25	0.25	0.25
Mean	73.5	84.5				
Sum				1514	1142	2062

$$r = \frac{1514}{\sqrt{1142}\sqrt{2062}} \approx 0,987$$



Regression Analysis

Examples:

- Can advertising spend explain changes in sales?
- Can market share be attributed to the size of the sales department?
- Is consumers' perception of quality influenced by their perception of price?



Regression Analysis

The **regression analysis** is a powerful and flexible tool for analyzing associative relationships between a metric **dependent** variable and one or more **independent** variables.

Enables you to:

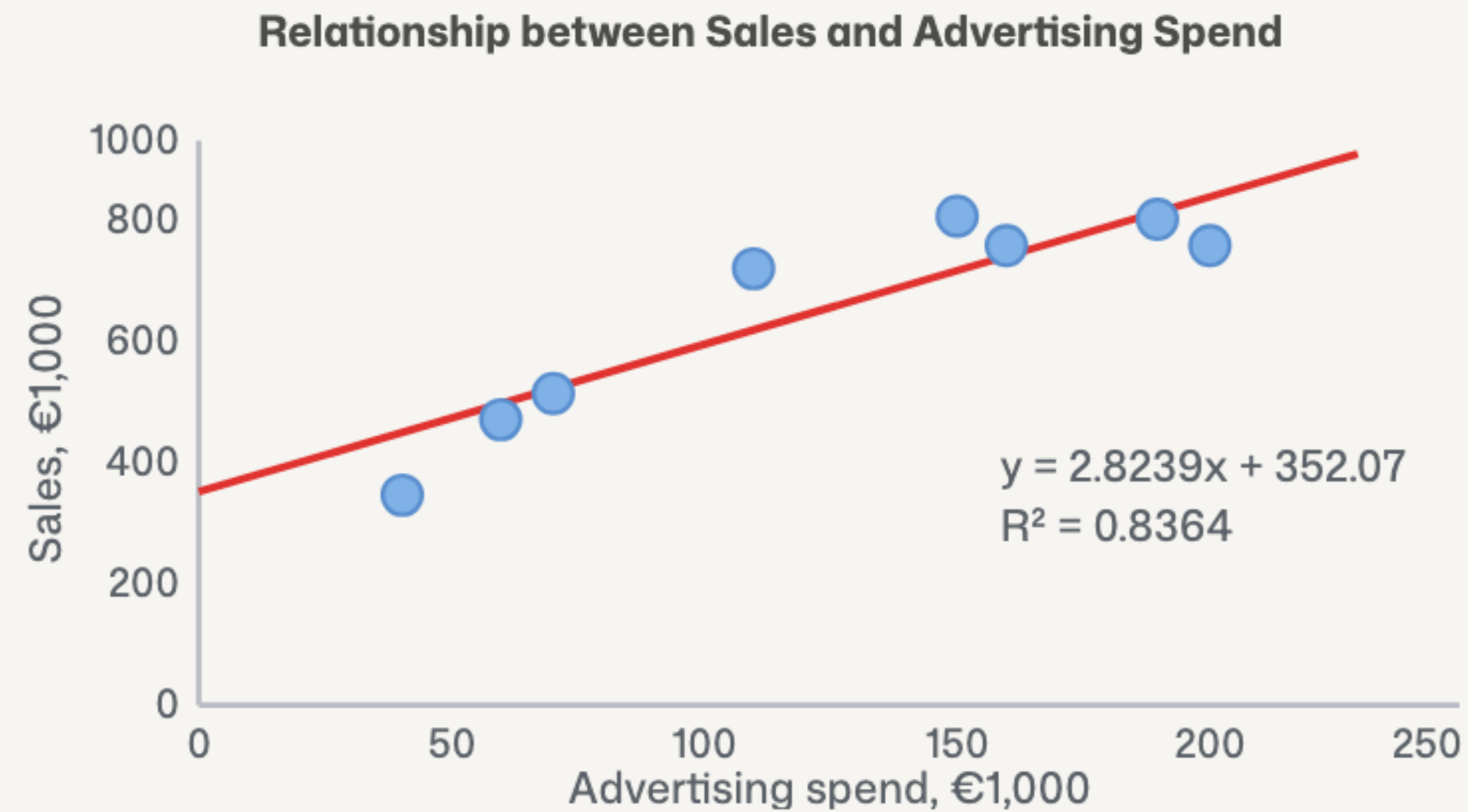
- determine the existence of the relationship,
- quantify the strength of the relationship,
- derive a mathematical model (formula) of the relationship,
- predict values of the dependent variable,
- account for the influence of other independent variables.

Regression Analysis

How many product units will we sell if we spend **€85.000** on advertising?

Advertising spend, €1.000	Sales, €1.000
40	377
60	507
70	555
110	779
150	869
160	818
190	862
200	817

Collected Data



- Advertising spend explains **83.6%** of the variance in sales.
- Every additional euro invested in advertising brings **€2.82** of additional sales.
- **€85.000** of advertising spend results in $2.8239 \cdot 85 + 352.07 = \mathbf{592.1}$ (thousand €) of sales.

6

Advanced Techniques of Market Analysis

Some Useful Concepts

- 6.1 **Conjoint Analysis**
- 6.2 **Market Simulations**
- 6.3 **Segmentation**
- 6.4 **Perceptual Positioning Maps**

6

Advanced Techniques of Market Analysis

Some Useful Concepts

- 6.1 Conjoint Analysis**
- 6.2 Market Simulations
- 6.3 Segmentation
- 6.4 Perceptual Positioning Maps

Conjoint Analysis

Conjoint Analysis

The **conjoint analysis** is a set of techniques used in market research to analyze **attribute-based preferences** of consumers – i.e. to determine how important different product features and feature levels are to consumers.

What should our new product look like?

Distinctive features

- Evaluation of **holistic objects**
- **Decomposition** of preferences

Widely used in

- Segmentation
- Development of new products
- Pricing



Conjoint Analysis

Please indicate how important the following PC features are for your next decision to purchase a PC?

	not at all important				extremely important
Manufacturer/Brand	☐	☐	☐	☐	☐
Processor performance, GHz	☐	☐	☐	☐	☐
Memory, GB	☐	☐	☐	☐	☐
Display size	☐	☐	☐	☐	☐
Price	☐	☐	☐	☐	☐

Conjoint Analysis

Expectation We **expect** respondents to reveal clear differences in importance to us.

Question: "Please indicate how important the following PC features are for your next purchase decision?"

	not at all important			extremely important		
Manufacturer/Brand	☐	☒	☐	☐	☐	☐
Processor performance, GHz	☐	☐	☒	☐	☐	☐
Memory, GB	☐	☐	☒	☐	☐	☐
Display size	☐	☐	☐	☒	☐	☐
Price	☐	☐	☐	☐	☒	☐

Conjoint Analysis

Reality In reality, practically everything is important.

to respondents

	Question: "Please indicate how important the following PC features are for your next purchase decision?"				important extremely
	not at all				
important	☐	☐	☐	☐	☒
Manufacturer/Brand	☐	☐	☐	☐	☒
Processor performance, GHz	☐	☐	☐	☐	☒
Memory, GB	☐	☐	☐	☐	☒
Display size	☐	☐	☐	☐	☒

Conjoint Analysis

The relative importance of individual product attributes can be measured more accurately when respondents evaluate them as a **single stimulus** (**CON**sidered **JOINT**ly) than when each attribute is evaluated in isolation:

- Many consumers are unable to determine the relative importance of individual product attributes.
- Individual attributes are perceived differently in isolation than in the combination that makes up a whole product.
- Constructing the preferred combination of attributes strains respondents' cognitive abilities – “all attributes are important.”
- Social desirability and a sharply defined self-image motivate respondents to give some attributes great weight even when they play no role (e.g. environmental friendliness, wealth, reduced importance of price).
- Some respondents deliberately try to manipulate the results by giving “advantageous” answers (e.g. overstating the importance of price).

Conjoint Analysis

If you had to buy a laptop today and these were the only alternatives available on the market – which alternative would you choose?

MacBook Air	Dell XPS 14	ASUS Zenbook	None
M3 chip (8 cores) 16 GB memory 13.6" display € 1.299,-	Core Ultra 7 32 GB memory 14.5" display € 1.899,-	Ryzen 7 (8 cores) 16 GB memory 14" display € 999,-	If these were the only alternatives, I would postpone the purchase.
☐	☐	☐	☐

6

Advanced Techniques of Market Analysis

A Brief Overview of Some Useful Concepts

- 6.1 Conjoint Analysis
- 6.2 Market Simulations**
- 6.3 Segmentation
- 6.4 Perceptual Positioning Maps

Market Simulations

Given the following preference values – **which product should we offer on the market?**

	Product alternatives		
	● Blue	● Red	● Yellow
Respondent #1	50	40	10
Respondent #2	0	65	75
Respondent #3	40	30	20
Mean	30	45	35

Market Simulations

Given the following preference values – **which product should we offer on the market?**

	Product alternatives		
	● Blue	● Red	● Yellow
Respondent #1	50	40	10
Respondent #2	0	65	75
Respondent #3	40	30	20
Mean	30	45	35

„**Red**“ has the highest average preference value.

Market Simulations

Given the following preference values – **which product should we offer on the market?**

	Product alternatives			Choice
	● Blue	● Red	● Yellow	
Respondent #1	50	40	10	Blue
Respondent #2	0	65	75	Yellow
Respondent #3	40	30	20	Blue
Mean	30	45	35	Red

“Red” does have the highest average – **yet no one actually chooses “Red.”**

Market Simulations

Suppose: **80 %** of customers prefer **round** things, **20 %** prefer **square** ones. What kind of things should we bring to market?

☒ **80% round things**

☐ **20 %**

WITHOUT ADDITIONAL INFORMATION

The choice seems obvious – go where most customers are:



“round things”

the “fat” part of the market

WHAT IF ...

there are already **10 competitors** that **ALL** offer round things?



10 providers compete for the same 80% – the 20% “square” are **completely unoccupied**.

Only **in the context of competition** does it become clear: the smaller, unoccupied market can be more profitable – and that is exactly what a **market simulation** makes visible.

Why Market Simulations?

Market simulations reflect reality **better than purely data-driven models** – and deliver decisions that actually hold up in competition.

01 Closer to reality

- capture the idiosyncratic preferences of segments and individuals
- account for preferences and competing offerings on the market

02 Niches, not just mass

No compulsion to fixate on the “fat” part of the market – unoccupied segments can also yield good profit.

03 “Test laboratory”

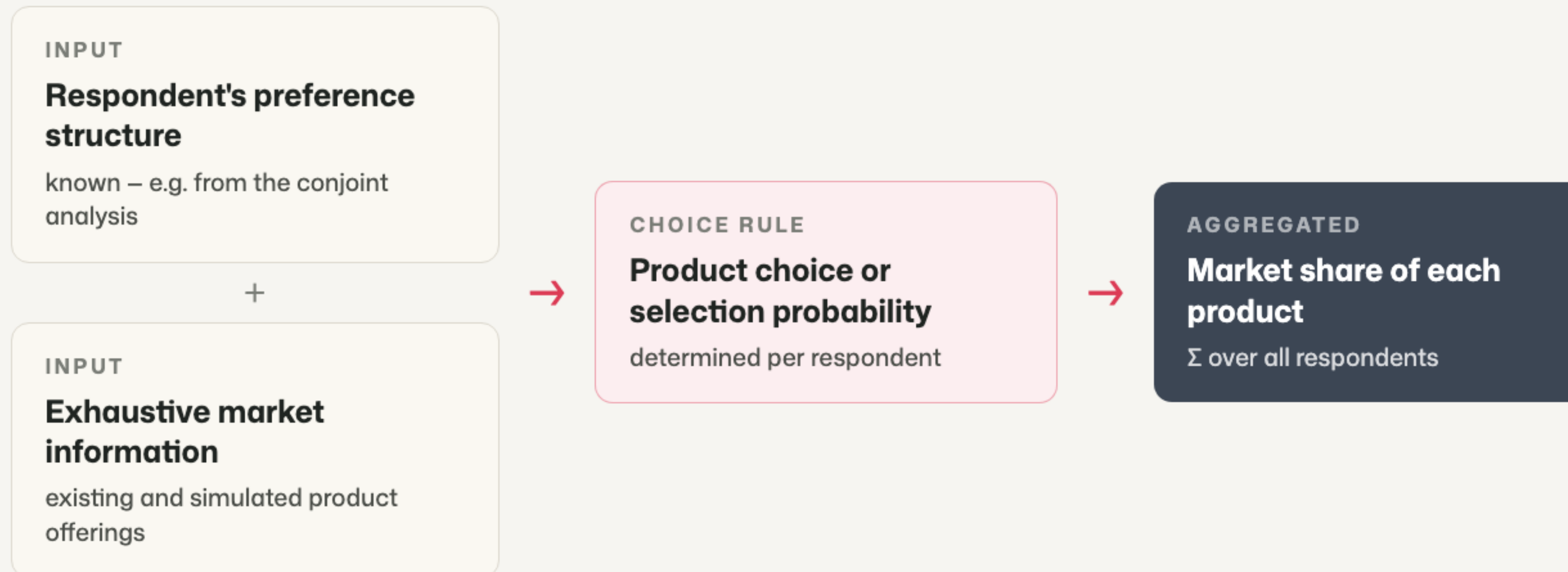
A multitude of real market opportunities can be played through risk-free – together with their possible outcomes.

04 Actionable for management

The results are easy to understand and can be translated directly into decisions.

What Do Market Simulations Do?

The basic process runs **individually for each respondent** – and is then aggregated into market shares:



What Do Market Simulations Do?

The product choice per respondent is determined by so-called **choice rules** – e.g.:

First-choice rule

First-Choice Rule

- The product with the **highest utility** is chosen.
- Selection probability = **100 %** for this product, **0 %** for all others.

BTL model

after Bradley · Terry · Luce

- Selection probability depends on the **relative utility share** in the market.
- Even products with low preference or utility value receive a **positive** probability.

$$\pi_h = \frac{U_h}{\sum_{h=1}^H U_h}$$

Logit rule

contrast-based

- Selection probability increases with growing **contrast** in product utility.
- Enables an **a-priori adjustment** of simulated to real market shares.

$$\pi_h(\alpha) = \frac{e^{\alpha \cdot U_h}}{\sum_{h=1}^H e^{\alpha \cdot U_h}}$$

6

Advanced Techniques of Market Analysis

Some Useful Concepts

- 6.1 Conjoint Analysis
- 6.2 Market Simulations
- 6.3 Segmentation**
- 6.4 Perceptual Positioning Maps

Market Segmentation

MARKET SEGMENTATION

Market segmentation refers to dividing the "relevant market" into groups of consumers who are **internally homogeneous** and **externally heterogeneous** – and forms the basis for differentiated market cultivation.



From a **mixed overall group**, internally homogeneous, mutually **distinct** segments emerge.

OBJECTIVE

Developing efficient **product differentiation strategies** – for the most optimal exploitation of the potential of individual segments.

Effective Market Segmentation

Six criteria determine the efficiency and economic viability of a segment solution:

1 Identifiability

Consumers can be identified on the basis of easily measurable variables.

2 Substantiality

Segments must be large enough to amortize investments.

3 Accessibility

Targeted communication or targeted use of the marketing mix is possible.

4 Stability

... over the period of planning, implementation and effect of segment-specific measures.

5 Behavioral relevance

Uniform response to segment-specific measures (e.g. price change).

6 Actionability

Meaningful and helpful in formulating the marketing mix.

Typology of Segmentation Bases

	General	Product-specific
Observable	• Cultural characteristics • Geographic characteristics • Demographic characteristics • Socioeconomic characteristics	• User status & usage situation • Usage frequency & intensity • Brand loyalty & allegiance
Not observable	• Psychographic characteristics • Values • Personality & lifestyle	• Benefit perceptions / product benefit • Attitudes & perception • Preferences, motives & intentions

Benefit Segmentation

The **benefit** that consumers expect from products is considered one of the most relevant segmentation bases of all:

“

... The benefits which people are seeking in consuming a given product are the basic reasons for the existence of true market segments. Thus, benefit is the most relevant segmentation base.

– Haley, 1968

“

... Benefit is one of the most popular bases of segmentation – for the purpose of market understanding, positioning, the development of new product concepts as well as advertising and distribution strategies. All of this owing to its actionability.

– Wind, 1978

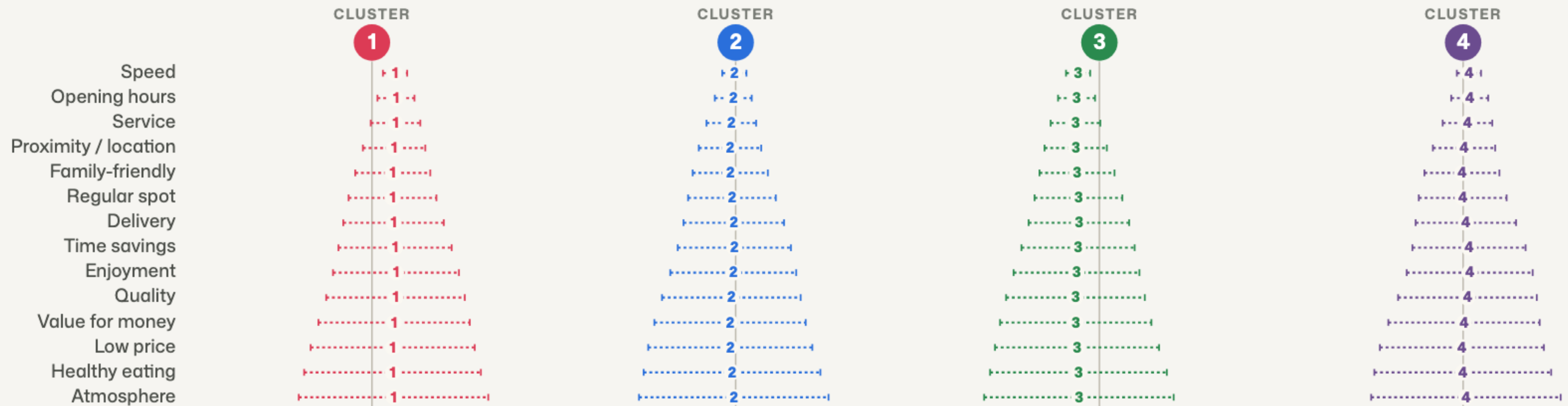
Evaluation of Segmentation Bases

Type of base	Identifi-ability	Substan-tiality	Accessi-bility	Stability	Action-ability	Behavioral relevance
1 · General, observable	++	++	++	++	-	-
2 · Specific, observable						
Purchase	+	++	-	+	-	+
Usage	+	++	+	+	-	+
3 · General, not observable						
Personality	±	-	±	±	-	-
Lifestyle	±	-	±	±	-	-
Psychographic characteristics	±	-	±	±	-	-
4 · Specific, not observable						
Psychographic characteristics	±	+	-	-	++	±
Perception	±	+	-	-	+	-
Benefit or benefit perceptions	+	+	-	+	++	++
Intentions	+	+	-	±	-	++

Highlighted – the **most promising bases** in practice: In combination they best cover the segmentation criteria.

++ very suitable + suitable ± conditional - unsuitable

How do the segments differ?



Trunk = overall mean across all respondents

Branch (dashed) = cluster mean $\pm$ 1 standard deviation

Order top → bottom: **F-ratio** descending (top best-separating)

Schematic illustration to demonstrate the principle

At the top are the characteristics with the highest **F-ratio** – small dispersion, clear position relative to the trunk: they **separate** the segments best. Toward the **bottom** the dispersion grows, the clusters overlap – separating power decreases.

Benefit Segmentation Paradox

When segmenting by expected benefit, one must clearly distinguish between **two types of variables**:

TYPE (I) · SEPARATING

Discriminating variables

Important for separating the sample into **internally homogeneous segments**.

TYPE (II) · UNDERSTANDING

Driver variables

Important because they represent the **benefit or characteristics** that respondents demand most within each segment.

It is tempting to assume that the two are the same. **Sometimes yes – but most of the time not:** The "drivers" often do not vary between segments at all and have no discriminating power (e.g. price, quality).

6

Advanced Techniques of Market Analysis

Some Useful Concepts

- 6.1 Conjoint Analysis
- 6.2 Market Simulations
- 6.3 Segmentation
- 6.4 Perceptual Positioning Maps**

Positioning

POSITIONING

Positioning aligns all marketing activities with the **preference structures** of potential customers – taking into account the **competing products**.

Objective: to design the company's offerings so that the **actual characteristics as perceived by customers** are brought into alignment with the **desired target characteristics**.



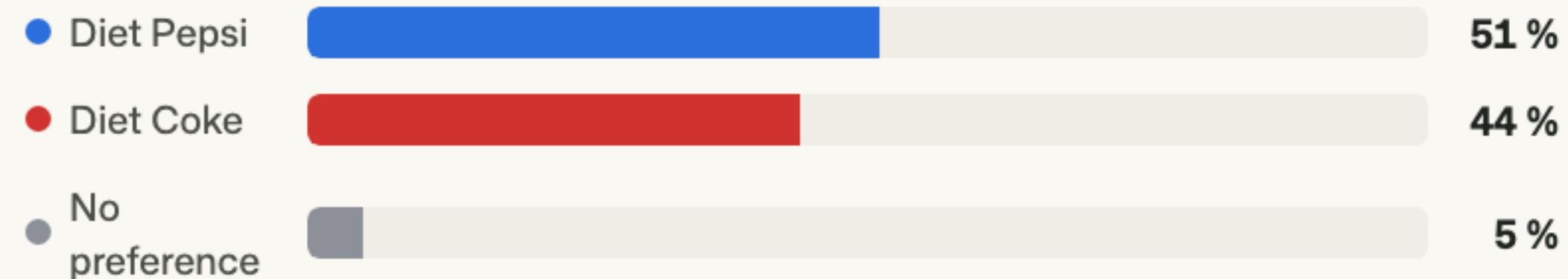
Like the red apple among the green ones: a **clearly distinguishable** position.

The Role of Perception

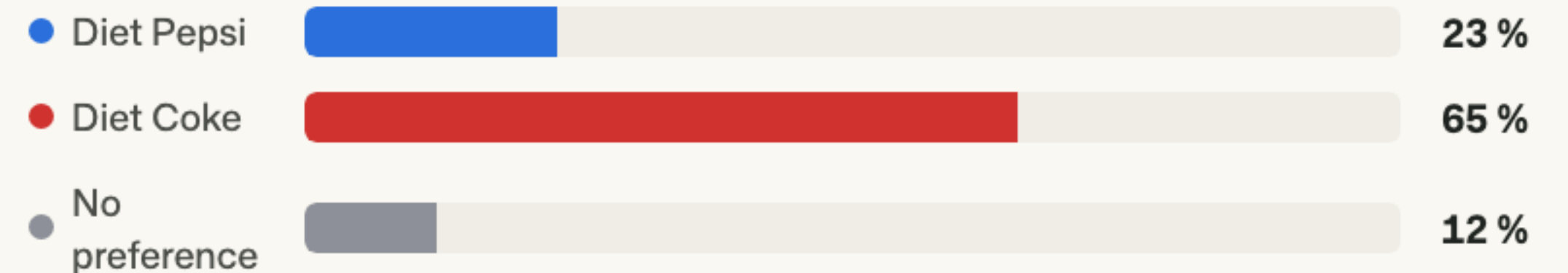
The same respondents, the same colas – only the knowledge of the brand changes. The preference **flips**:



Blind test Respondents don't know which cola they are drinking

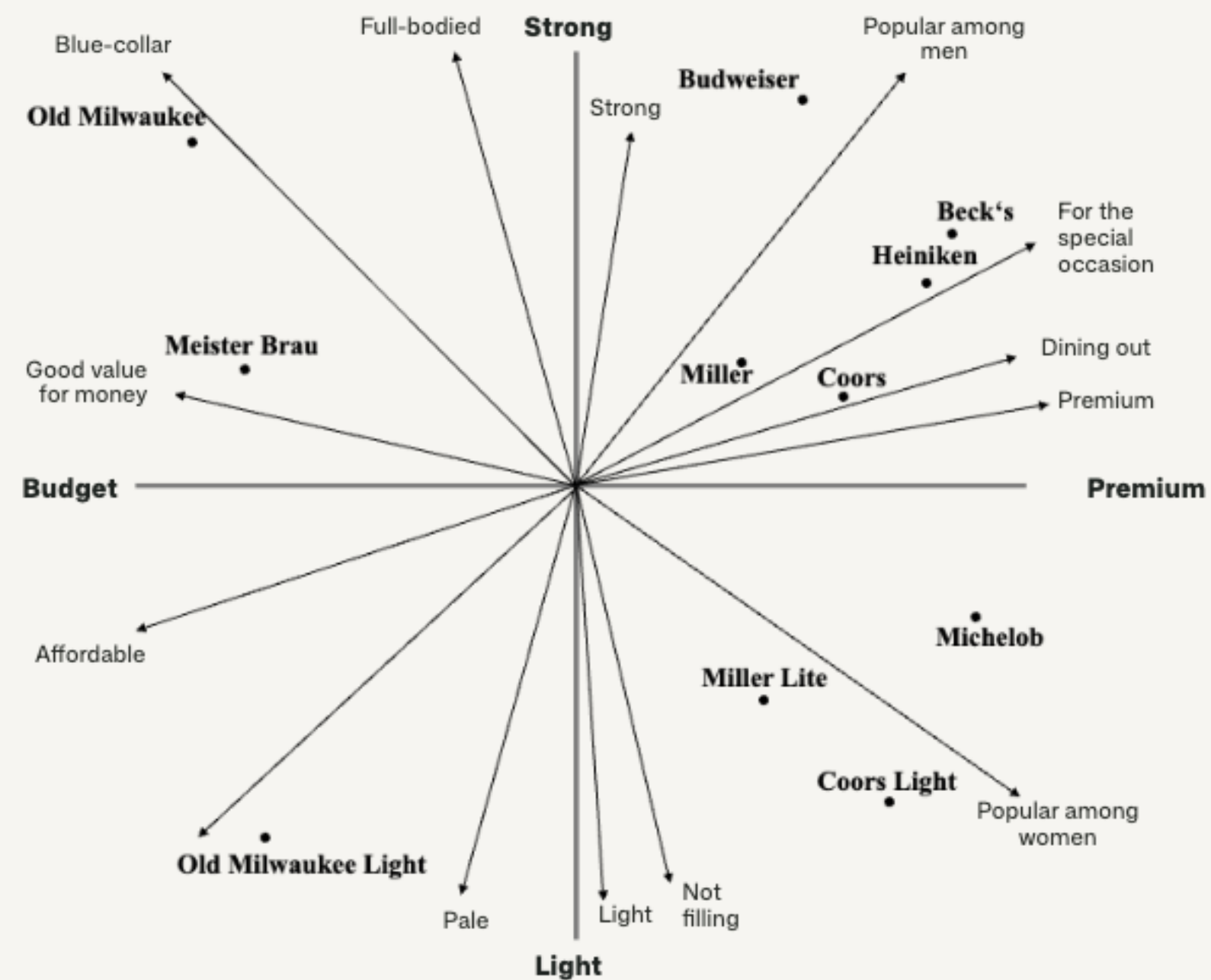


"Open" test Respondents know which cola they are drinking



Blind, **Pepsi** wins – with brand knowledge, **Coke** wins. It's not the product that decides, but **perception**.

Perceptual Positioning Maps



Perceptual positioning maps depict the positions of competing products, brands or companies in a "virtual" attribute space – the way consumers perceive the entire product category.

Axes

The latent product attributes that **differentiate best**.

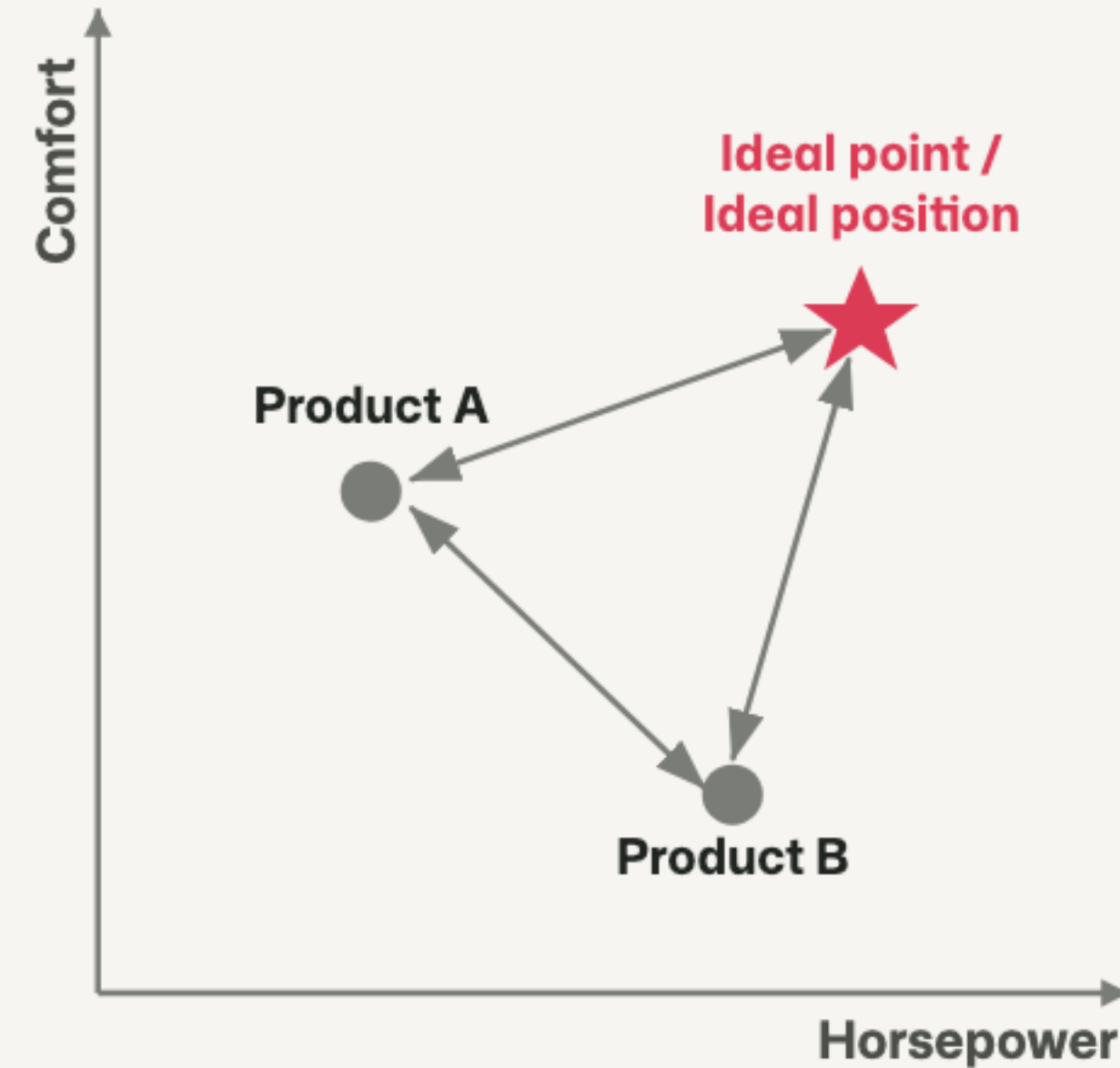
Vectors

Indicate the **direction and strength** of perceived product attributes.

Distances

Between two alternatives correspond to the degree of their perceived **(dis)similarity**.

Objective of Positioning

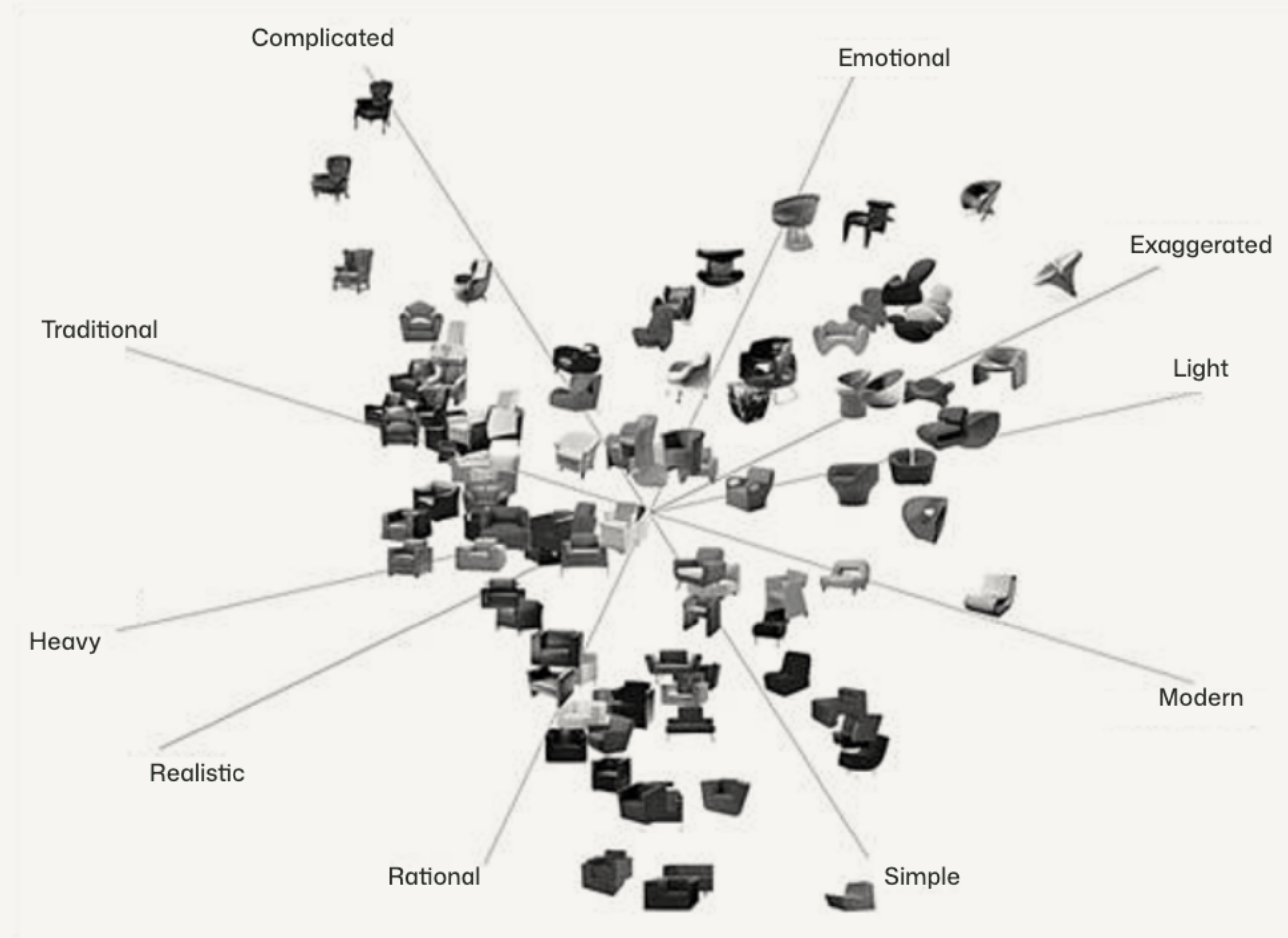


The objective of positioning is to occupy such a position in consumers' perception that is:

- as close to the ideal point as possible, and
- as far from the competition as possible.

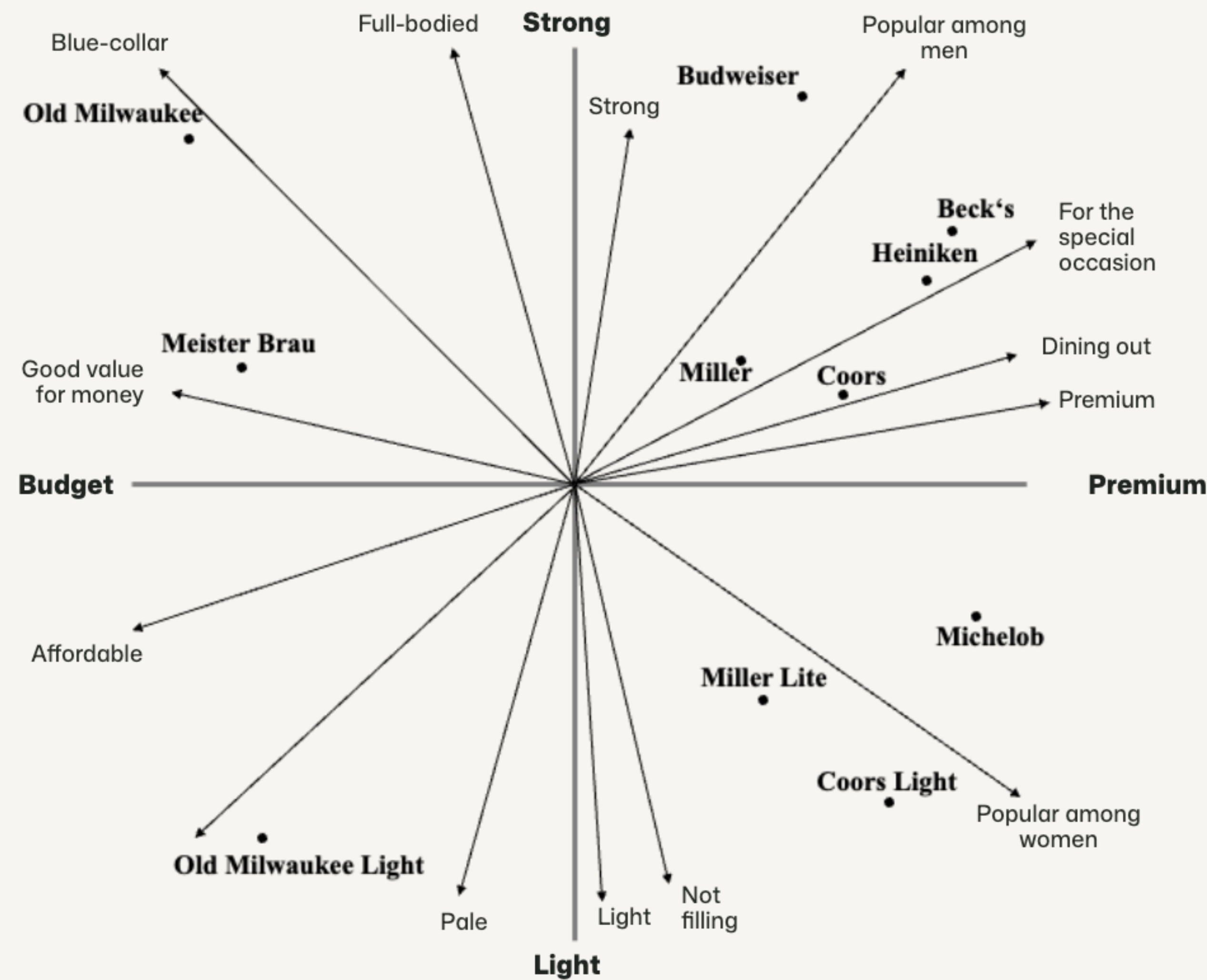
Each axis is a **latent attribute**; each product occupies a perceived position.

Perceptual map of armchair designs



Source: Chuang & Chen (2008), International Journal of Design

Perceptual Positioning Maps

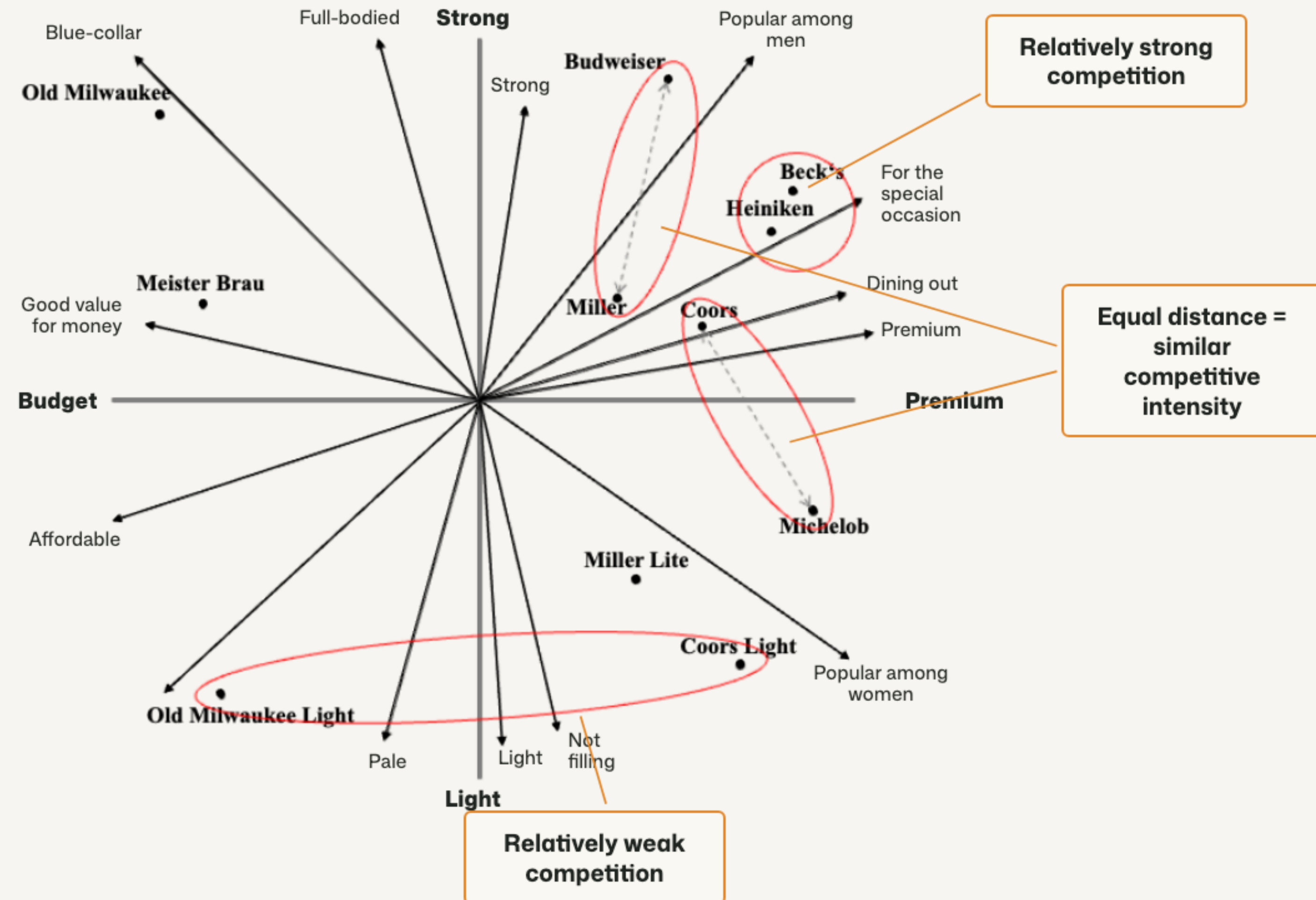


Source: Moore / Pessemier (1993), p. 145

Perceptual Positioning Maps

Competitive intensity

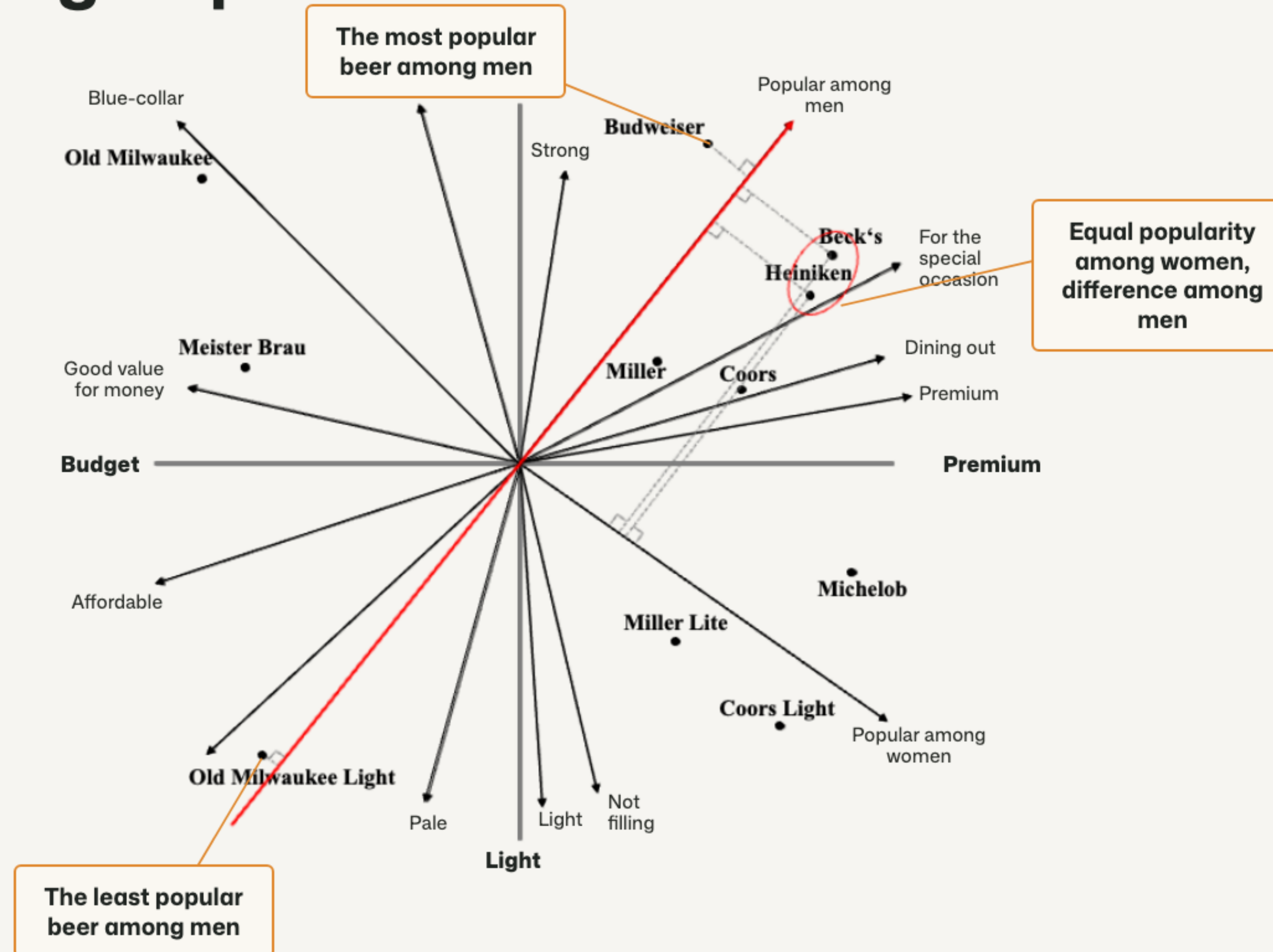
The closer the brands are, the more similar they are in consumers' perception – the stronger ("more direct") the competition.



Perceptual Positioning Maps

Attribute vectors

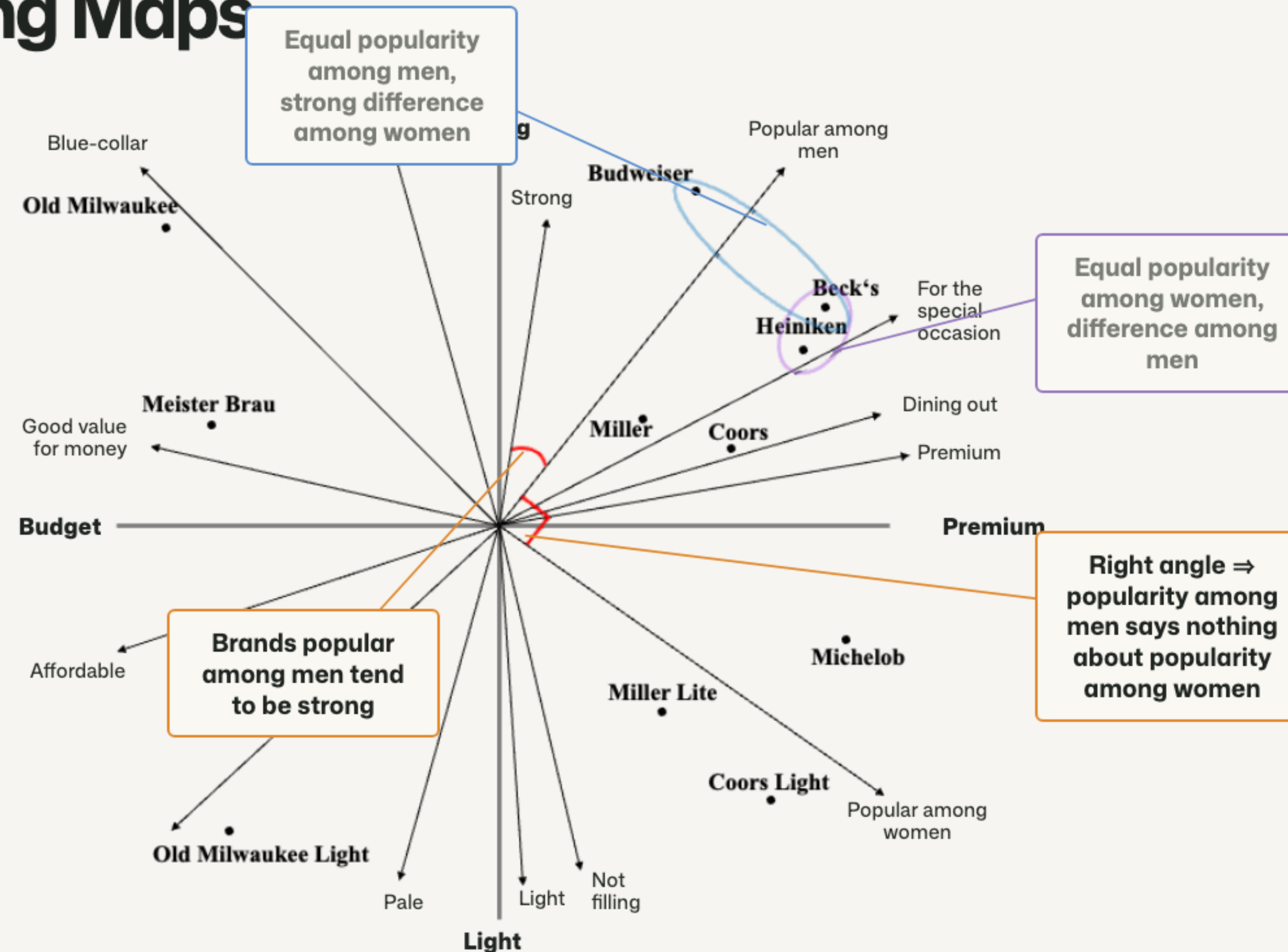
Consumers' perception of brands: The farther a brand is from the origin **along an attribute vector**, the more strongly this attribute is expressed in it.



Perceptual Positioning Maps

Relationships between attributes

The **smaller the angle** between attribute vectors, the higher their pairwise correlation.



Perceptual Positioning Maps

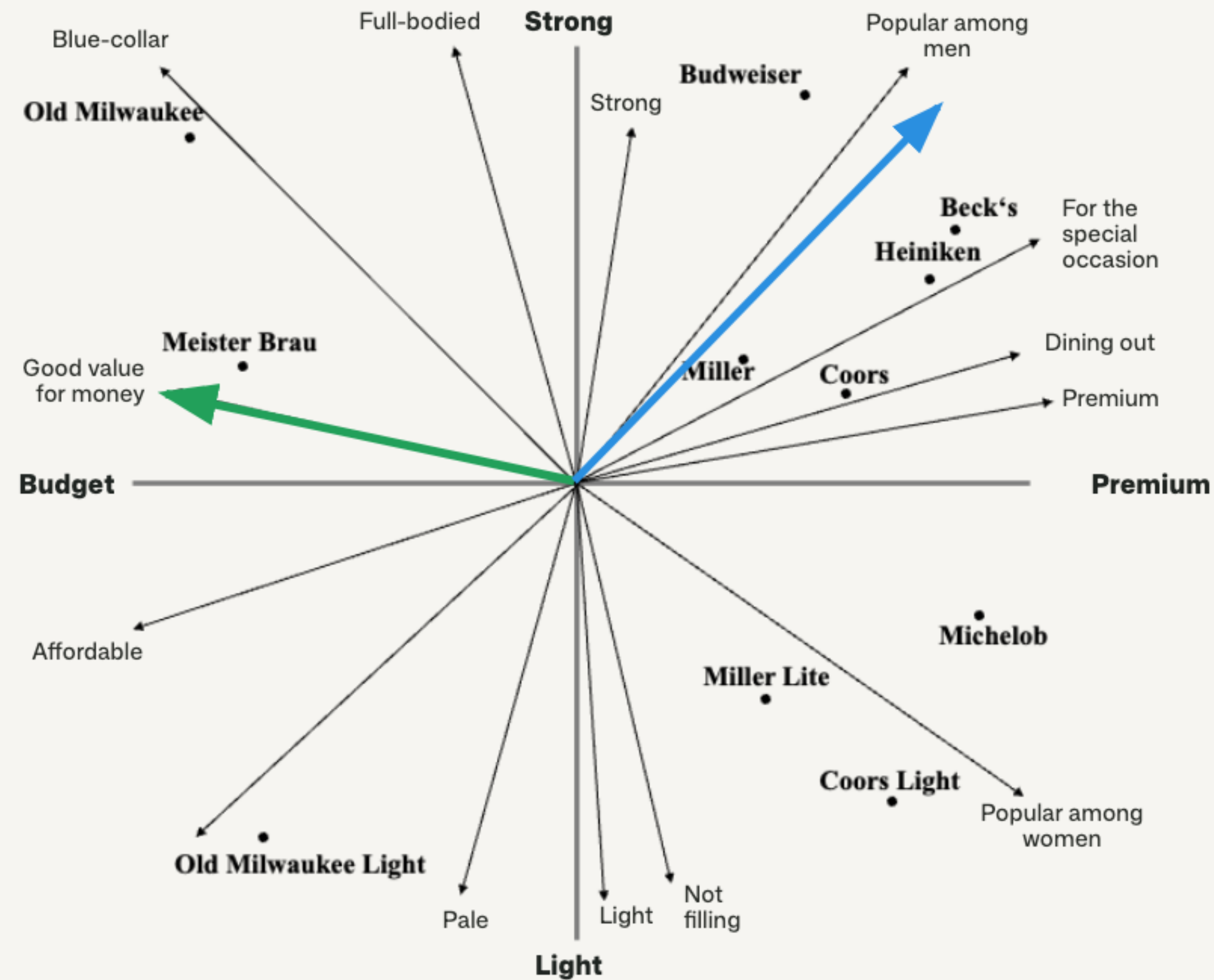
The length of the attribute vector indicates its degree of differentiation

The **longer the vector**, the more strongly this attribute differentiates between the beers.

→ Good value for money

→ Popular among men

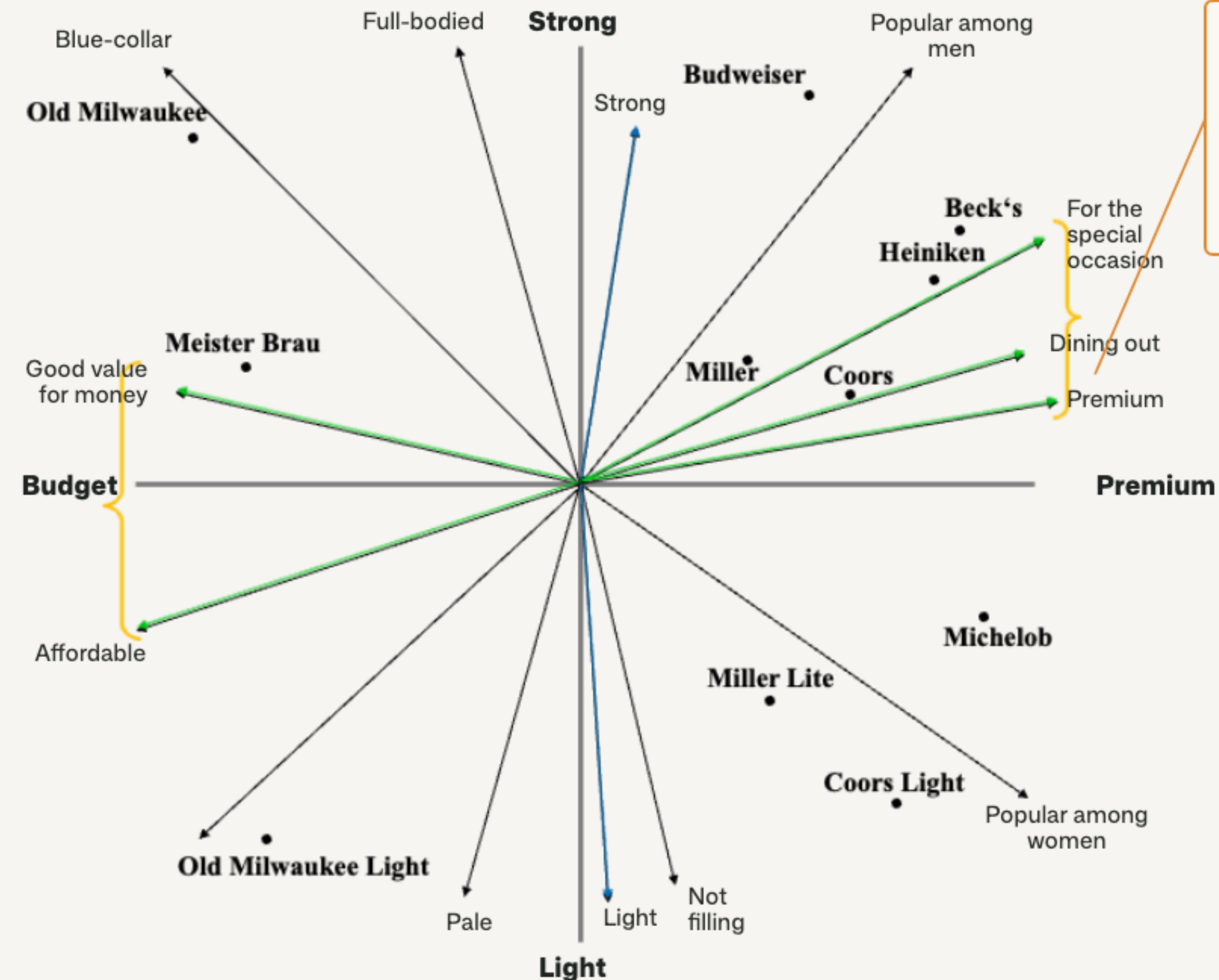
Consumers can **"Popular among men"** differentiate the brands better on the dimension **"Good value for money"**.



Perceptual Positioning Maps

Axes differentiate most strongly

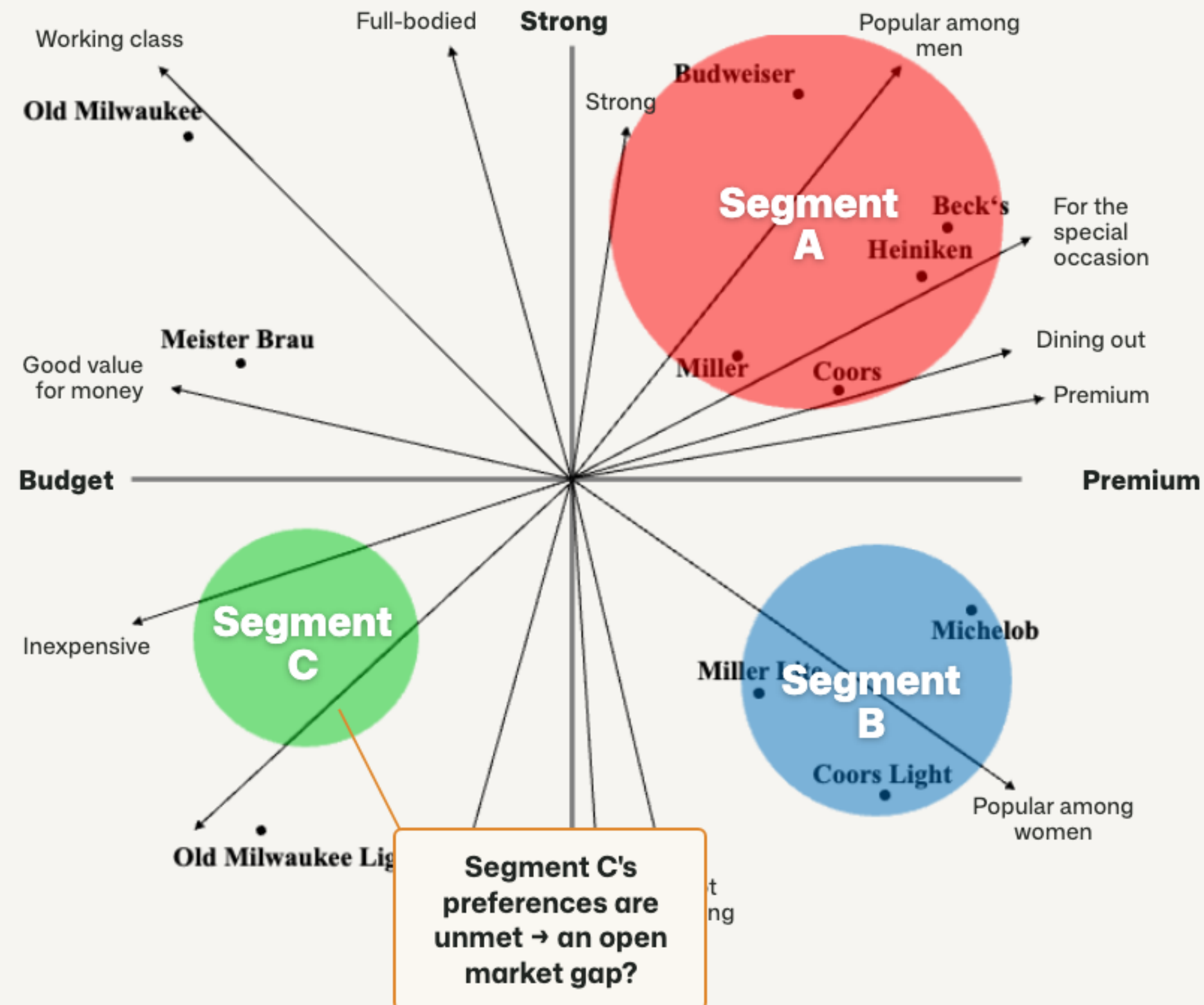
Axes are "**virtual**" **attribute vectors** that differentiate most strongly. Their label is usually derived from the neighboring vectors.



Perceptual Positioning Maps

Perceptual Positioning Maps and Segmentation

Straightforward assessment of **segment potential** and attractiveness; straightforward formulation of positioning strategies, statements and advertising campaigns.



Possible themes for advertising campaigns:

Segment A: Strong beer for strong men.

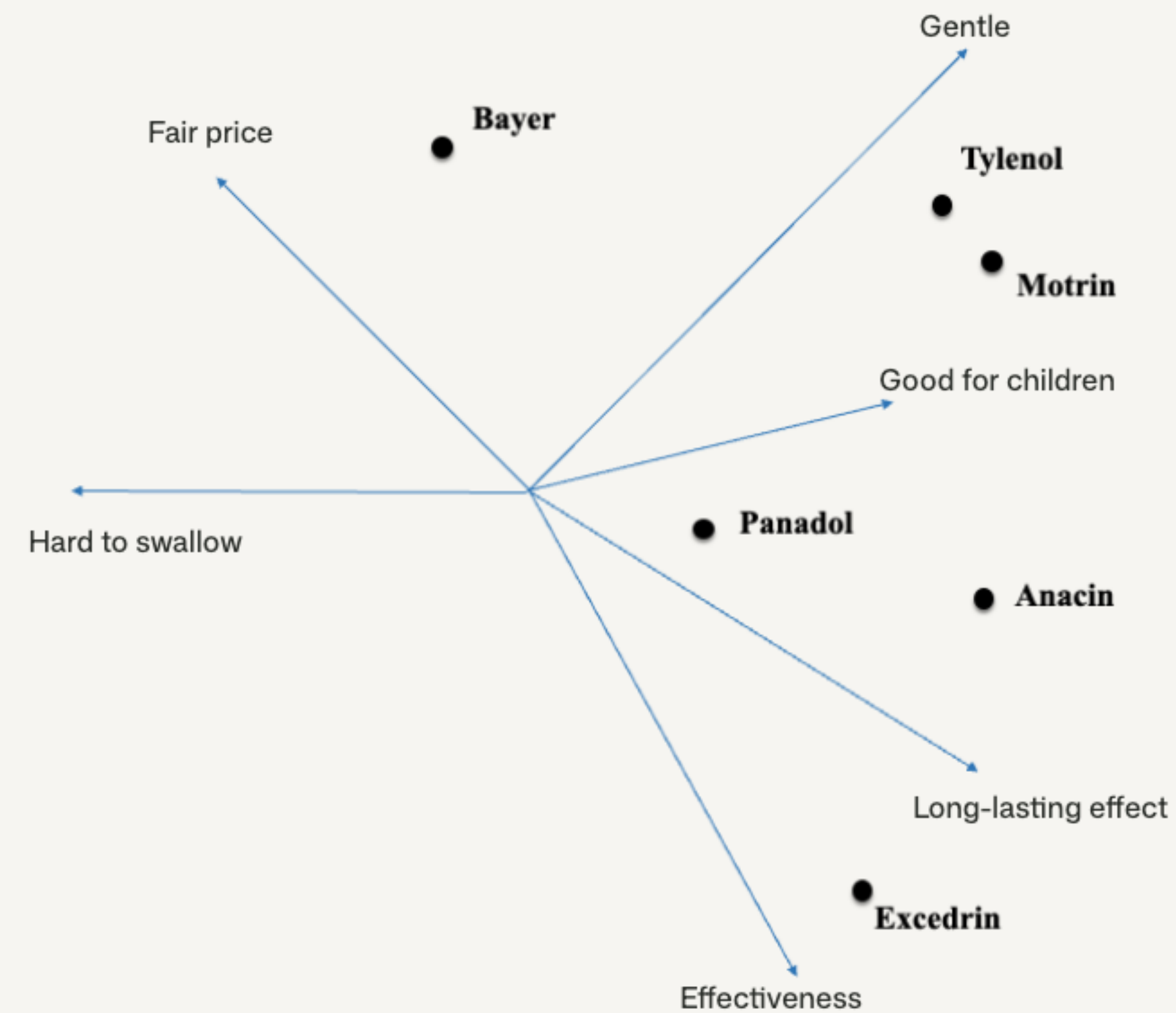
Segment B: Real ladies drink premium light beer on special occasions.

Segment C: Light lager – good beer at a good price.

Example: Perception Map of Pain Relievers

Test yourself:

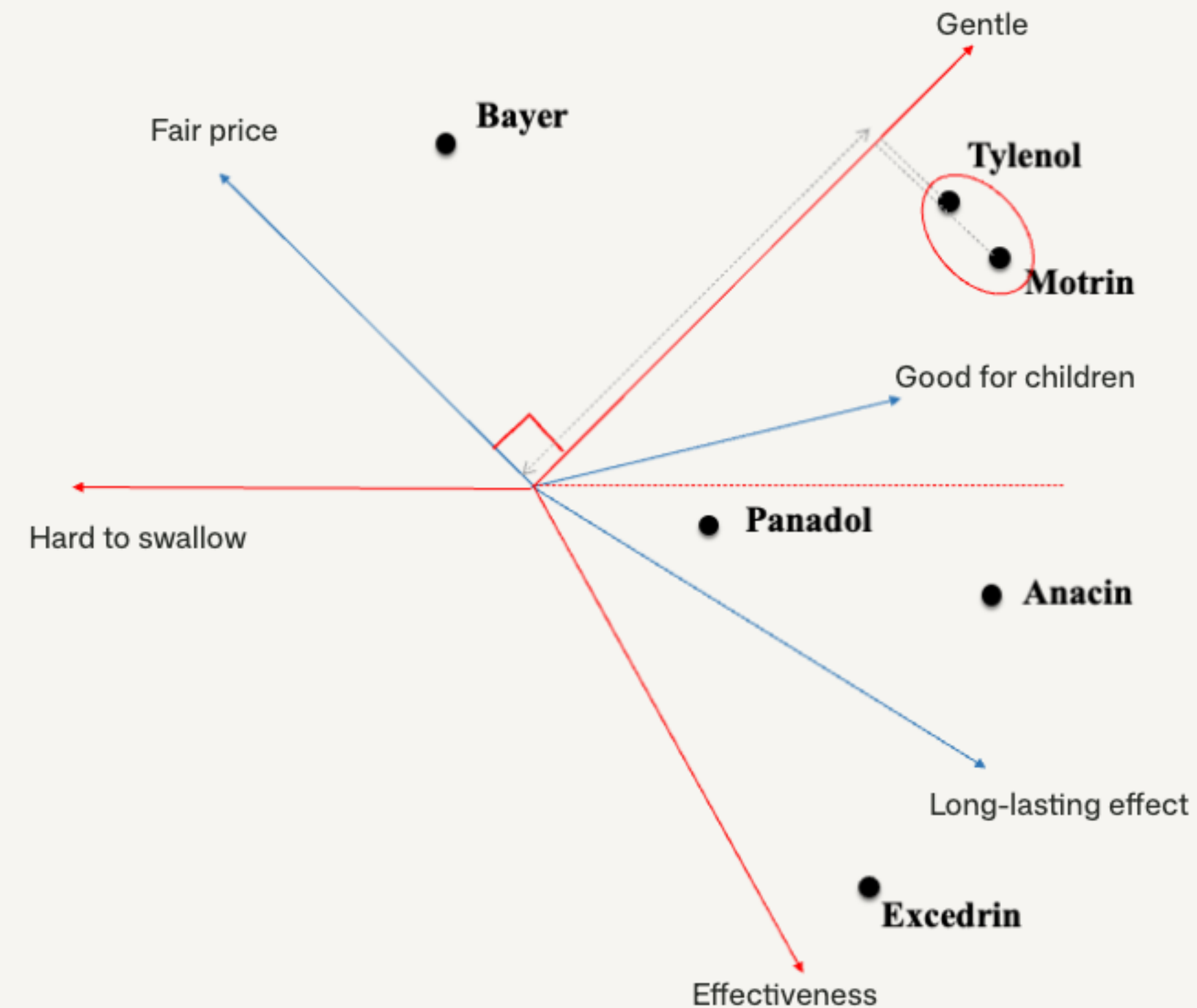
- Which medications face the strongest competition?
- How can these medications best be described from a marketing perspective in terms of just **one** of the attributes shown?
- Which **two** attributes best communicate a medication's advantages?
- Can a company justify a higher price on the grounds that its medication is gentler than all the others?
- Which attribute (other than “Good for children”) should a manufacturer of children's medications optimize first?



Example: Perception Map of Pain Relievers

Test yourself:

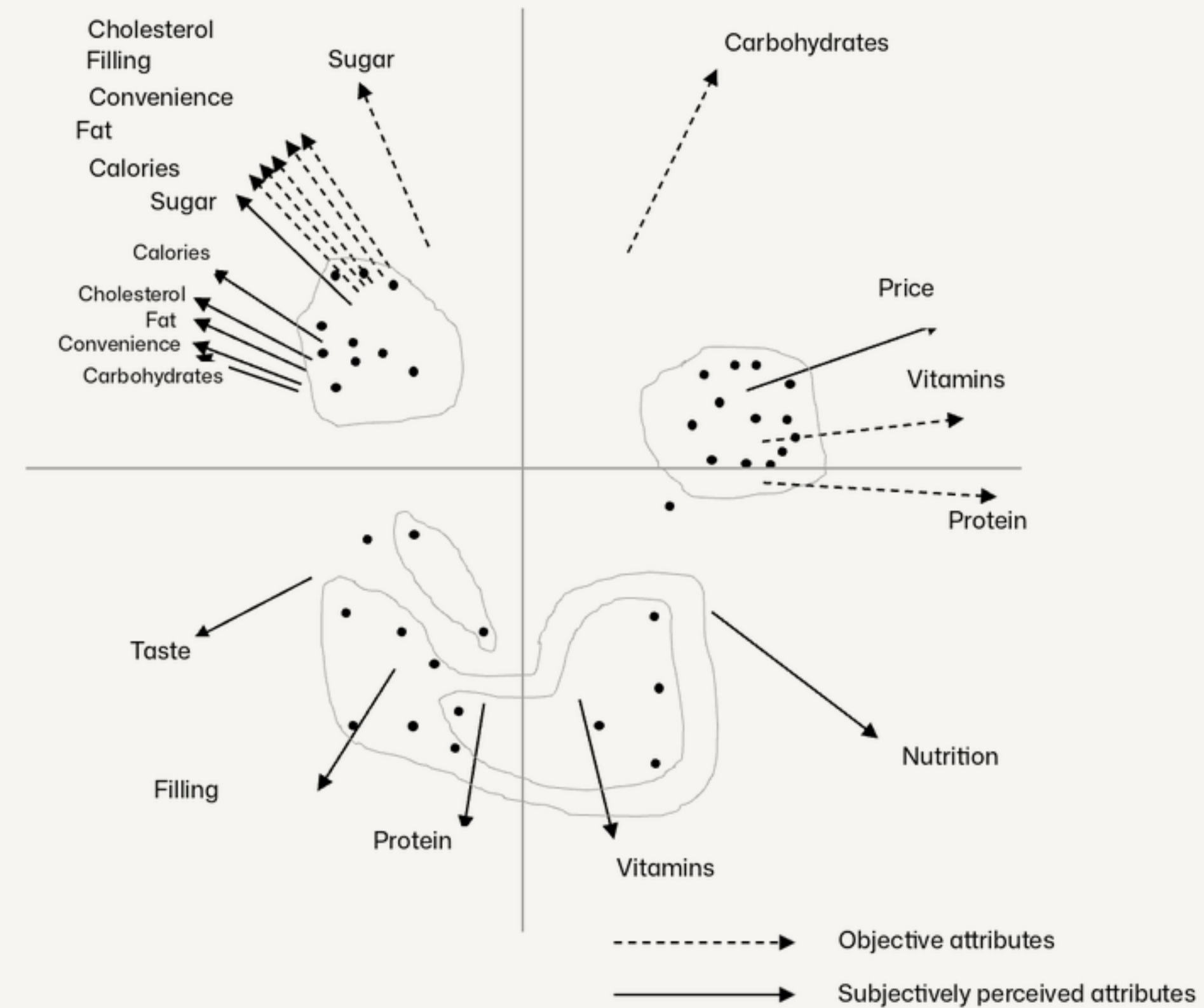
- Which medications face the strongest competition?
Tylenol and Motrin (closest to each other)
- One attribute for the best description?
“Gentle” – farthest from the origin along this vector
- Which two attributes best communicate the advantages?
“Gentle” and “Effectiveness” (the longest vectors)
- Higher price because gentler than all the others?
No! In perception, these attributes are independent of one another.
- Which attribute (other than “Good for children”) to optimize first?
“Hard to swallow” – smallest angle to “Good for children”.



On Importance of Perception in Positioning

A curious but typical case

40 foods: **subjective perception** vs. **objective attributes**.

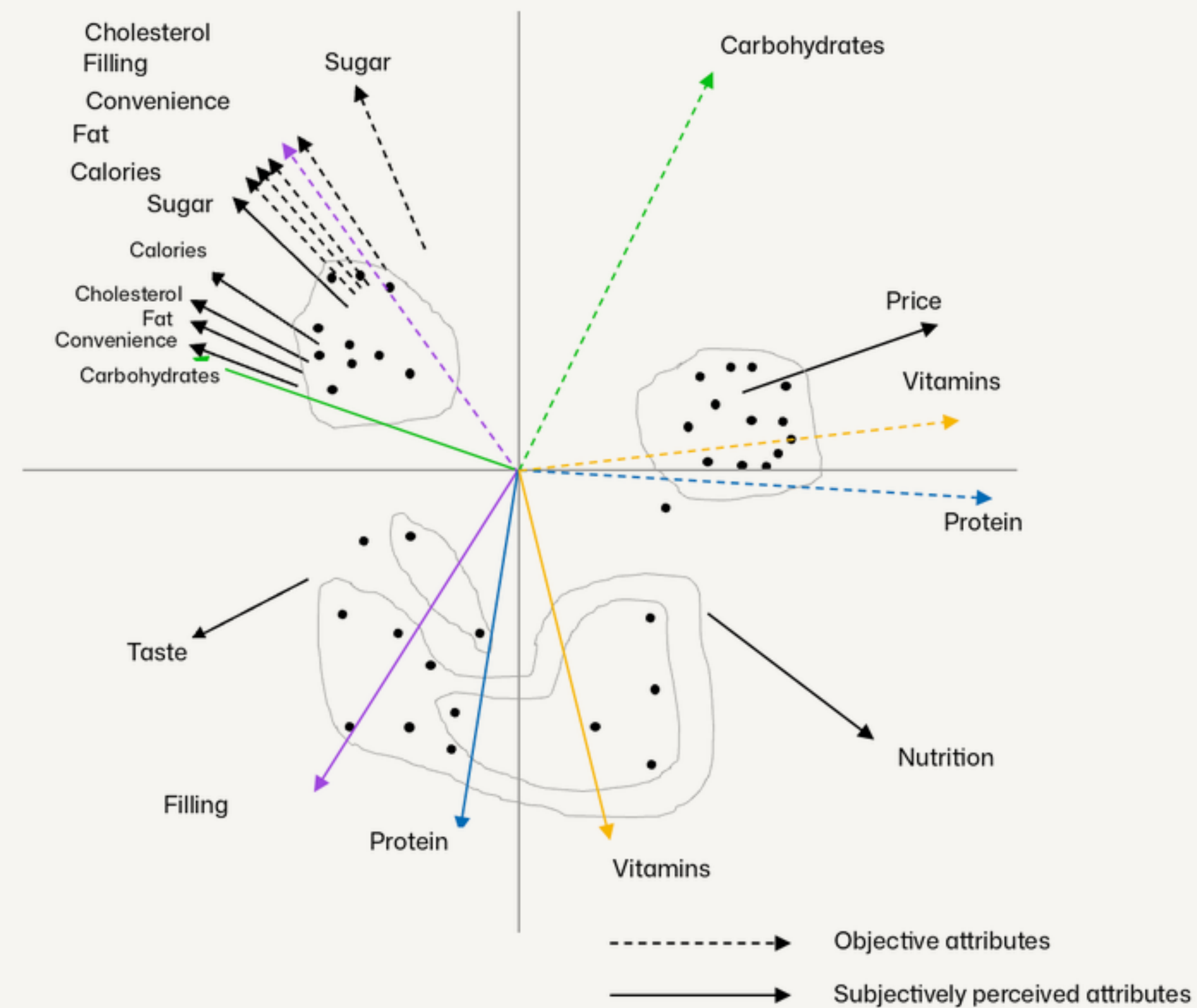


On Importance of Perception in Positioning

A curious but typical case

40 foods: **subjective perception** vs. **objective attributes**.

The perception or assessment of many food attributes often has **nothing or very little** to do with the actual content of these attributes.



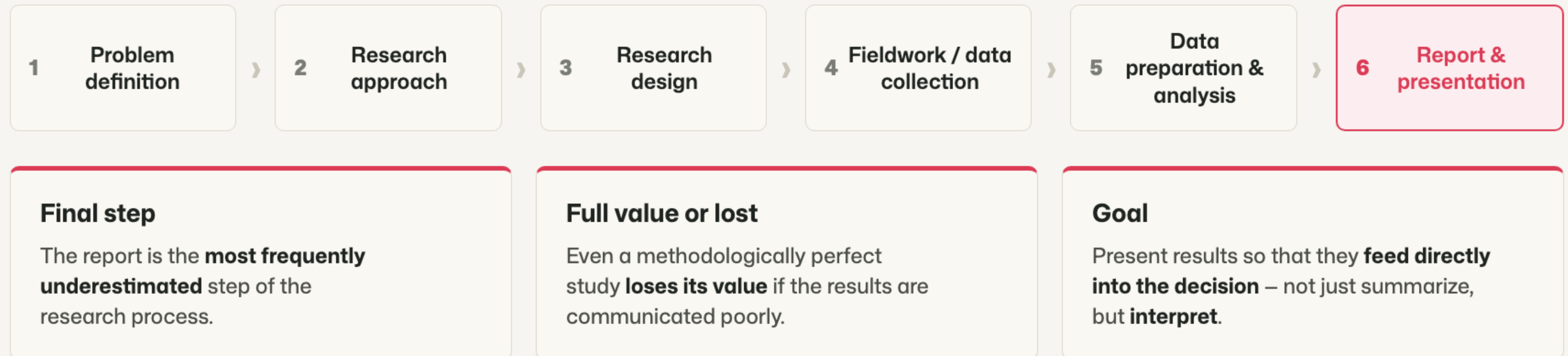
7

Reporting Results

Reporting & Presentation – Stage 6 of the research process

Managers should **easily understand** the report, **trust** the results and know **which actions** they should take.

The final stage of the research process



Why the report and presentation matter



Tangible product

After the project ends, often only the written report remains – it serves as a **historical record** of the entire study.



Basis for decision

The management decision rests on the report. A **weak finish** devalues all the preceding research.



Sole point of contact

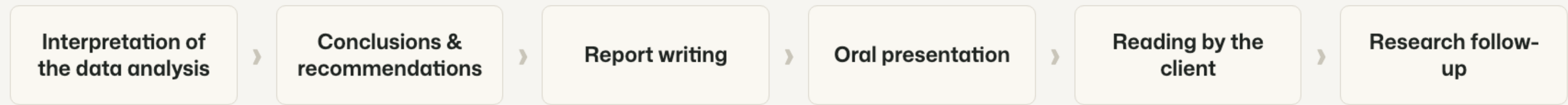
Many managers experience only the report and the talk – and **judge by them** the quality of the entire project.



Basis for follow-up assignments

The decision for future research or the same service provider hinges on the **perceived benefit**.

From data result to follow-up



Interpret rather than summarize

The results should serve directly as **input into the decision-making**; where useful, conclusions are drawn and **actionable recommendations** given.

Align in advance

Discuss the key findings, conclusions and recommendations **before writing** with the decision-makers – this ensures fit, acceptance and delivery dates.

The typical structure – eleven elements

Front matter

- Cover letter
- Title page
- Table of contents
- Executive Summary

Main body

- Problem definition
- Approach & research design
- Data analysis
- Results

Closing

- Conclusions & recommendations
- Limitations & caveats

+ **STANDALONE Appendix** – Authorization, questionnaire, sampling details, technical tables.

What comes before the report itself

Cover Letter

Delivers the report, sums up the project experience (**without results**) and points to necessary follow-up steps.

Title Page

Title in a **manager's tone** rather than "research-speak"; details on researcher and client, date.

Table of Contents

Main and sub-headings with page numbers; then lists of **tables, figures, appendices**.

Executive Summary

Often the only part management reads.

Briefly describes **problem, approach and design** and devotes one section to the key **results, conclusions and recommendations**. Written **last** – only once the entire report is finished.

The substantive core

1

Problem Definition

Background, conversations with decision-makers and experts, then a clear **management and research question**.

2

Approach & Design

Theoretical foundations, models, hypotheses; methods presented graphically and non-technically, **details in the appendix**.

3

Data Analysis

Justify the analysis plan and techniques, explain them in **simple terms with examples**.

4

Results

The longest part. Structured by form of analysis, data collection method or objectives – tied to the information needs.

5

Conclusions & Recommendations

Don't just summarize: **interpret** and – where possible – derive actionable recommendations.

6

Limitations & Appendix

Name limitations **in a balanced way**, without undermining confidence; appendix with authorization, questionnaire, sample.

Principles of good writing

Reader-oriented

Write for the **decision-makers**; avoid jargon, put technical terms in the appendix.

Easy to follow

Logical structure, headings, **short clear sentences**; have outsiders proofread it.

Professional

Careful design; vary typography – but only as far as it **supports understanding**.

Objective

Present design, results and conclusions accurately, **do not tailor them to the expectations** of management.

Text ↔ Graphics

Reinforce key information with tables/figures – and bring figures **to life with quotes**.

Concise

Leave out everything unnecessary – but **never at the expense of completeness**.



"The readers of your reports are busy people – hardly anyone can balance report, coffee and dictionary all at once."

Guidelines for tables

- **Number & Title** – a unique Arabic number, a short descriptive title, referenceable in the text.
- **Arrangement of Data** – order by the most important aspect: time, magnitude or alphabetically.
- **Unit of Measurement** – state the base clearly (column vs. row percentages, sample size).
- **Reading Aids** – lines, shading or white space guide the eye across the row.
- **Headings, Stubs, Footnotes** – column heads, left margin column, explanatory footnotes.
- **Source** – for secondary data, name the data source.

EXAMPLE · TABLE 25.1 – GLOBALCASH

EMU Impact	Total	Single	Dual	Multiple
Existing relationships	46 %	41 %	45 %	48 %
Fewer banks (Eurozone)	33 %	31 %	30 %	35 %
One main bank coordinates	33 %	43 %	39 %	29 %
Fewer banks per country	22 %	15 %	17 %	25 %

Stub = left margin column · **Heading** = column head · % as column percentages

Source: GlobalCash-Europe98, Statistical Report for Europe, p. 137

Chart types at a glance



Maps

Geographic and positioning maps show locations, customers, competitors – the basis of **geodemographics**.



Pie Chart

Simple relative frequencies. **Not** for time series or multiple variables.



Line Chart

Connects data points – ideal for **trends over time**; multiple series comparable.



Bar / Histogram

Shows absolute/relative magnitudes and differences. **Histogram** = vertical frequencies.



Stacked / grouped

Few data points, to represent **differences between groups** qualitatively.



Diagrams & Flowcharts

Represent **process steps** or the linking of qualitative ideas.

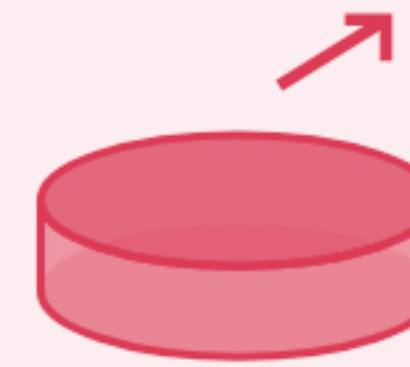
Two rules you won't forget



PIE CHART

max. 7 segments

Suitable for simple relative frequencies – **not** for time series or relationships between multiple variables.



CAUTION

No 3D charts

3D **distorts relative magnitudes** and confuses the audience. Programs offer many 3D options – hardly any presents data clearly and without distortion.

The talk shapes the first impression

Preparation is everything

Script/outline **after the report**, rehearse repeatedly.

Tailor to the audience

Know the background, interest and stake of the listeners.

Visual media

Flip chart, projector, software – never lose sight of the **message**.

Body language & voice

Vary gestures, eye contact, volume and pace; **a strong close**.



"My paradigm is the guitar." Technique is only a tool – a good guitar alone does not carry the performance. **You do.**

"Predictability precedes boredom."

Research Follow-up

1

Support the client

Explain technical parts, help with implementation, discuss follow-up projects and **integrate results into the MIS/DSS**.

2

Evaluate the project

While it's still fresh, ask critically: "Could the project have been carried out **more effectively or efficiently**?"

Trust as the key

The quality of the **personal interaction** between manager and researcher shapes the perceived quality of the report itself. Trust influences relationship quality, engagement, retention – and ultimately **how strongly the market research is actually used**.

Reports across countries and languages

Multiple versions

For management in different countries and languages, **separate, reader-specific versions** – comparable in content, possibly different in format.

Cultural sensitivity

When presenting, observe cultural norms – **humor is not appropriate everywhere**. Adapt recommendations to be country-specific where needed.

Practical relevance

Language versions maintained in parallel (e.g. **DE / RU / KZ**) are not a translation problem but a **reporting problem**: field-by-field localization, comparable metrics, consistent terms.

Integrity in interpretation and reporting



33%

of the market researchers surveyed name questions of **research integrity** as their most difficult ethical problem.

TYPICAL VIOLATIONS

- Ignoring relevant data
- Compromising the research design
- Deliberately misusing statistics
- Falsifying numbers or altering results
- Reinterpreting results to favor a particular view
- Withholding information

Resist the temptation: Shaping ambiguous findings into a "coherent, well-formed story" is satisfying – but **unethical**. Maintain objectivity, even when nothing significant emerges.

From "Push" to "Pull"

BACK THEN

- Printed report, "push"
- Password-protected intranet reports
- First multimedia reports on the web
- Searchable, retrievable worldwide

TODAY

- Interactive dashboards, "pull"
- Real-time data & live filters
- Crosstabs and weighting on demand
- Linked reports, rules for robustness



The key takeaways

1

Last step, full leverage

The report and presentation determine the **perceived value** of the entire study.

2

Interpret, don't just recite

Prepare results as **decision input** – with actionable recommendations.

3

Readers first

Avoid jargon, structure clearly, let **text and visualization** reinforce each other.

4

Visualization with discipline

≤ 7 pie segments, be careful with 3D, label tables cleanly.

5

People before technology

The talk lives from the **presenter** – not from the slide.

6

Integrity & follow-up

Stay objective, follow up with the client, evaluate your own project.

About the Author



Dr. Paul Marx was a Professor of Marketing at the University of Siegen, where he conducted research particularly on e-commerce, preference measurement, new media, big data, and recommender systems. He studied aero- and hydrodynamics as well as management at Novosibirsk State Technical University (Russia) and subsequently held senior positions in marketing. In 2000 he moved to Germany, deepened his studies in economics at the University of Hannover, and founded the online service for online surveys **eQuestionnaire** – today **QUESTIONSTAR**. Paul earned his doctorate at the Bauhaus University of Weimar and published his research in leading international journals, including the **Journal of Marketing**.

Motaev Marx Motaev GbR

Vahrenwalder Str. 253 · 30179 Hannover · Germany **T** +49 511 89 86 15 34 **E** info@questionstar.com

References & License

REFERENCES

- **Backhaus, K., Erichson, B., Plinke, W., Weiber, R.** (2015): "Multivariate Analysemethoden: Eine anwendungsorientierte Einführung", Springer Gabler, 14th edition.
- **Brandt, D. R.** (1996): "Secure Customer Index", Maritz Research.
- **Bruner, G. C.** (2012): "Marketing Scales Handbook", Vol. 6.
- **Burke, Inc.** – burke.com.
- **Chuang, S.-C., Chen, H.-C.** (2008): International Journal of Design.
- **Malhotra, N. K.** (2020): "Marketing Research: An Applied Orientation", Prentice Hall, 6th edition.
- **Marx, P.** (1979–2026) – own international experience.
- **Moore, W. L., Pessemier, E. A.** (1993), p. 145.
- **Myers, J. H.** (1996): "Segmentation & Positioning for Strategic Marketing Decisions", South Western Educ. Pub.
- **Noelle-Neumann, E., Petersen, T.** (1998): "Alle, nicht jeder. Einführung in die Methoden der Demoskopie", Springer, p. 192.
- **Reichheld, F.** (2003): "The One Number You Need to Grow", Harvard Business Review.
- **Reichheld, F., Markey, R.** (2011): "The Ultimate Question 2.0", Harvard Business Review Press.
- **Sullivan III, M.** (2010): "Statistics: Informed Decisions Using Data", Pearson, 3rd edition.
- **Visocky O'Grady, J. & K.** (2009): "The Information Design Handbook", HOW Books.
- Course "Statistics I" of Elgin Community College.

LICENSE & DISCLAIMER

Disclaimer: The author and affiliated persons/organizations cannot be held responsible for any violation of license terms unless caused by their active conduct. Brand names and registered trademarks are the property of their respective owners and are named for descriptive purposes only. Errors excepted.

CC BY-NC-SA 3.0

This presentation is subject to the **Creative Commons Attribution-NonCommercial-ShareAlike license**, unless otherwise stated. Any use or distribution requires a reference to this presentation and the explicit mention of **Dr. Paul Marx** and **QUESTIONSTAR**.

©2002–2026 Dr. Paul Marx, QUESTIONSTAR. All rights reserved.