Usability Engineering

Data Analysis

Carsten Röcker
inIT - Institut für industrielle Informationstechnik
carsten.roecker@th-owl.de

Data Analysis

Main Statistical Methodologies

- Descriptive Statistics
  - aims at summarizing a sample
  - goal: organization and presentation of empirical data
- Explorative Statistics
  - similar to descriptive statistics
  - searches for unknown structures and relations
- Inferential Statistics
  - applies statistical procedures to test hypotheses

Descriptive Statistics

Two Sets of Properties

- Central Tendency
  - seeks to characterize the distribution's central or typical value
- Dispersion (or Variability)
  - characterizes the extent to which members of the sample depart from its center and each other

General Methods

- Tables
- Diagrams
- Indexes

Descriptive Statistics

Analyzing Empirical Data

- Counting Frequencies
- Drawing Diagrams
- Calculation of Mean Values
- Identification of Minimum and Maximum
- Calculation of Standard Deviation

Frequencies: Table

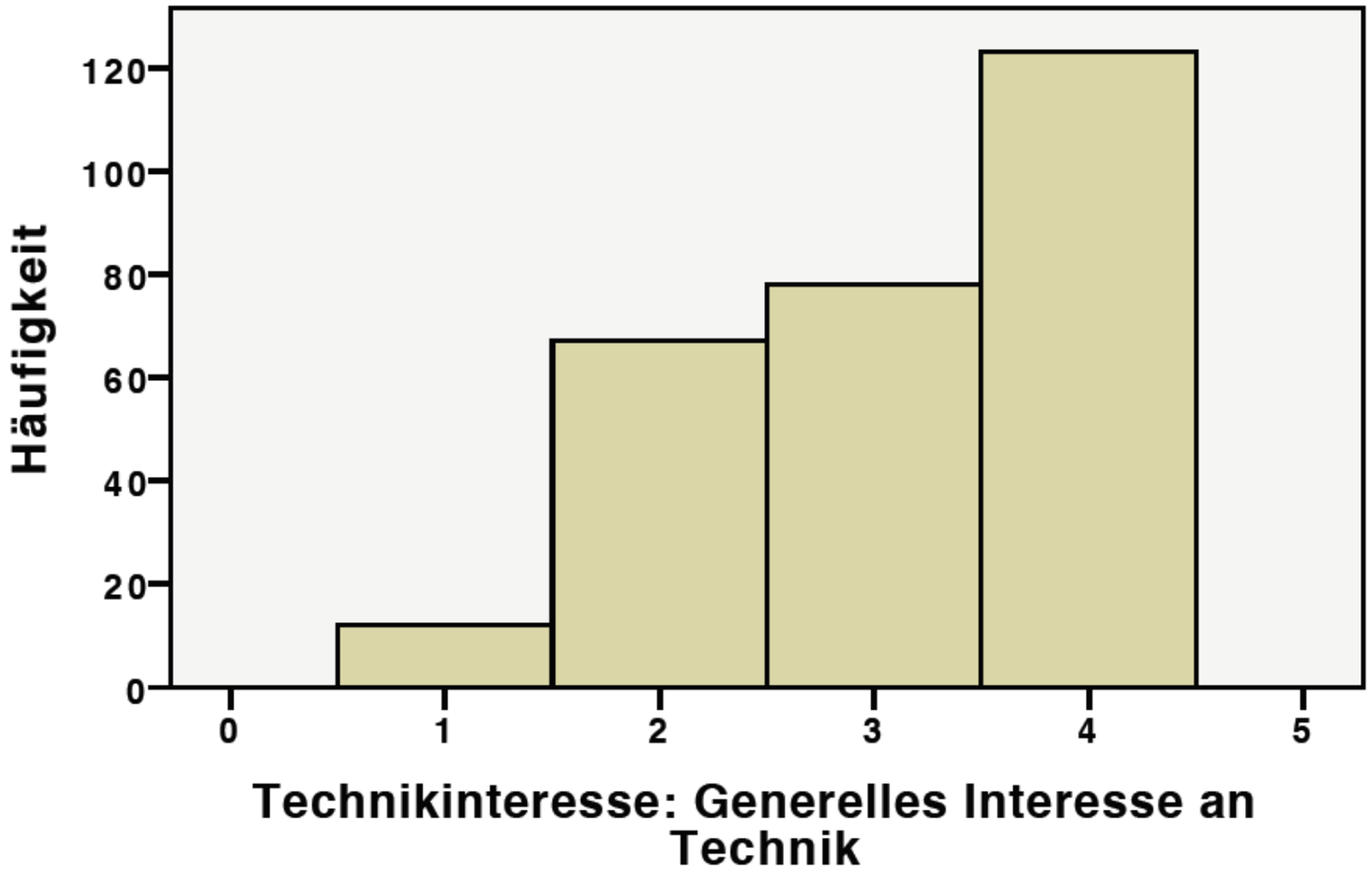
- counting of responses
- absolute frequency
- relative frequency

Technikinteresse: Generelles Interesse an Technik

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	12	4,3	4,3	4,3
	eher nein	67	23,9	23,9	28,2
	eher ja	78	27,9	27,9	56,1
	ja	123	43,9	43,9	100,0
	Gesamt	280	100,0	100,0	

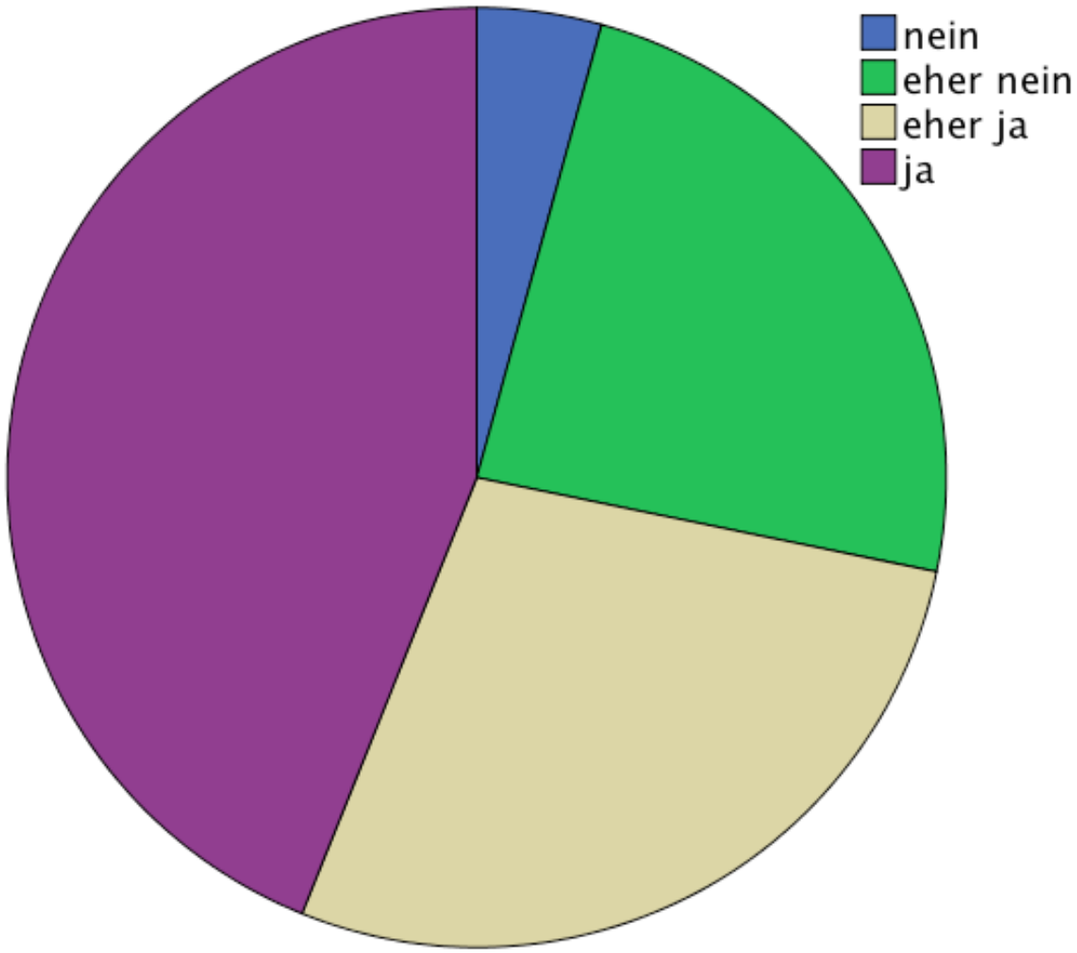
Frequencies: Histogramm

- Histogramm = graphical representation of the frequency distribution



Frequencies: Charts

- pie charts
- bar chart



Measures of Central Tendency

- describe the center of a distribution using a numerical value
- different measures
  - arithmetic mean
  - median
  - mode
- type of measure that should be used depends on the research question, the nature of data, and the scales of measurement

Arithmetic Mean

- the sum of all of the numbers in a sample divided by the number of items in that sample
- requires metric scale

$$\bar{x}_{\mathbf{arithm}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Median

- numerical value separating the higher half of a data sample from the lower half
  - i.e. 50% of the data are above and the other 50% are below the median
- in case there is no middle value, the median is calculated as the arithmetic mean of the two middle values
- advantage: no influence of extreme values (outliers)

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & n \text{ ungerade} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & n \text{ gerade.} \end{cases}$$

$$= \frac{1}{2}\left(x_{\lceil \frac{n}{2} \rceil} + x_{\lfloor \frac{n}{2}+1 \rfloor}\right) \qquad = \frac{1}{2}\left(\tilde{x}_u + \tilde{x}_o\right)$$

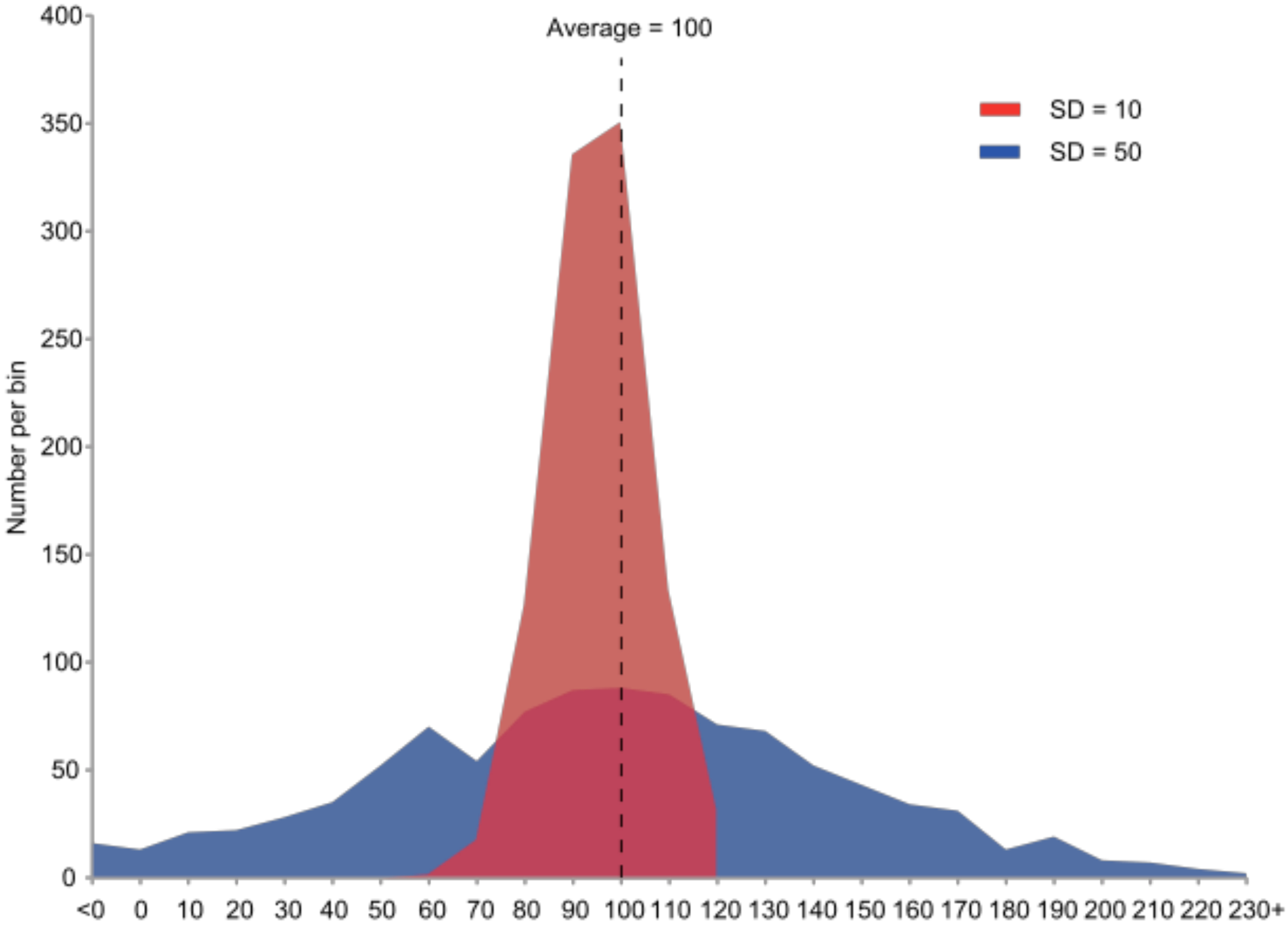
Mode

- value that appears most often in a set of data
- used for categorial variables

Standard Deviation

- measures the amount of variation from the average
  - low standard deviation: data points tend to be very close to the mean
  - high standard deviation: data points are spread out over a large range of values
- only usable for metric values

Standard Deviation: Example



Presentation of Results

- questionnaires use mostly ordinal scales
- numerical coding in SPSS „transforms“ data to interval scale
- numerical operations become possible
  - e.g., mean, standard deviation

Presentation of Results in Written Form

- „… the importance was rated on a 4-point scale (very important = 1, unimportant = 4) …“
- „… x was regarded as important (*M*=3.23, *SD*=1.23) …“