

VO 2 SWS / 4 ECTS
Einführung in die Statistik
für Linguist:innen
MASTER

SoSe 2024
Isabella Hager

Inhalte

1. Fragestellung, Operationalisierung, Stichprobendesign
2. Item-Formulierung und Fragebogenkonstruktion
3. Datenniveaus, Datenmatrix, Häufigkeitstabelle
4. Lage- und Streuungsmaße, Verteilungen, Normalverteilung
5. Kreuztabelle, Signifikanz, Chi-Quadrat-Test
6. Unterschiedstestungen: parameterfrei und parametrisch
7. Korrelation, Indexbildung

Note

1	91 bis 100
2	81 bis 90
3	71 bis 80
4	60 bis 70
5	0 bis 59

Take-Home-Exam (100 Punkte)

Zeitraumen: 3 Tage

Beginn: Mittwoch, 12. Juni 8 Uhr, Ende: Freitag, 14. Juni 20 Uhr.

Bearbeiten einer eigenen sprachwissenschaftlichen Hypothese.

<https://homepage.univie.ac.at/isabella.hager/sprachwissenschaft/master/pruefungsfragen.pdf>

Vorbereitung für die Prüfung:

Erarbeiten Sie die Prüfungsfragen mithilfe des Buches und der Folien.

Alle Infos sowie weitere Literatur finden Sie auf der HP zur Lehrveranstaltung.

Zur Ausarbeitung können Sie alle Unterlagen verwenden.



Achten Sie jedoch darauf, dass Sie keine Sätze oder Inhalte in Ihre Prüfungsarbeit kopieren. Formulieren Sie Ihre eigenen Sätze, entwickeln Sie eigene Gedanken. Ihre Prüfungsarbeit wird einer Plagiatsprüfung unterzogen.

Fragestellung

Operationalisierung

Empirische Sozialforschung = die systematische Erfassung und Deutung sozialer Wirklichkeit

empirisch

= erfahrungsgemäß: beobachtbar oder messbar; Theoretisch formulierte Annahmen werden anhand spezifischer Wirklichkeiten überprüft.

systematisch

= die Überprüfung erfolgt nach Regeln, ist somit nachvollziehbar (im Unterschied zur Alltagsbeobachtung).

Wissenschaftlichkeit: - **strukturiert:** geplanter Ablauf, festgelegtes Design

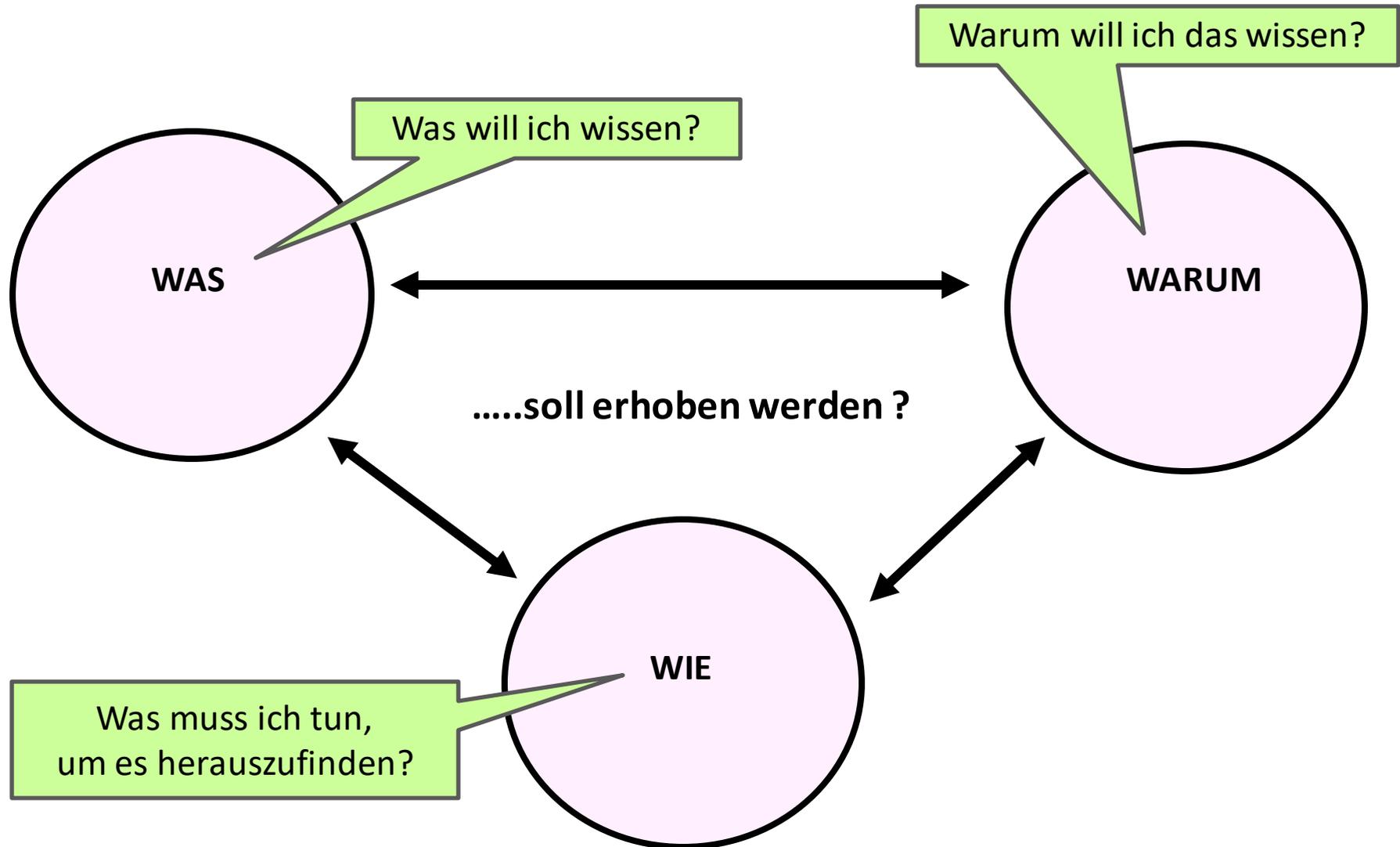
- **allgemein:** repräsentative Auswahl einer Stichprobe

- **kontrolliert selektiv:** festgelegte Kriterien über das, was beobachtet wird

soziale Wirklichkeit

= geschaffene Gegenstände, soziale Eigenschaften (Geschlecht, Alter, Einkommen...), Sprache, Sprachgebrauch, beobachtetes/mitgeteiltes Verhalten, mitgeteilte Einstellungen

Grundfragen empirischer Sozialforschung



hier, dass die reichen Kinder immer um ein bis zwei cm. die armen Kinder übertreffen, was uns den Schluss gestattet, dass unter unseren Kindern die wohlhabenden einen grösseren Kopfumfang haben als die armen.

Tabelle XI.

Kopfumfang.

Anthropologie der nicht-besitzenden Klassen, Alfredo Niceforo, 1910

KNABEN.

	REICH.	ARM.	
10 Jahre	528.0	523.3	5cm
11 „	533.0	524.8	↓
12 „	535.5	524.9	↓
13 „	536.4	528.6	↓
14 „	541.8	528.4	13cm

Eben dasselbe kann man für die Stirnhöhe sagen. Bei Betrachtung

ForscherInnen machen die Ergebnisse: Vorsicht bei der Lektüre von wissenschaftlichen Studien:

Welche **Fragestellung** ist relevant?

Wer bestimmt die **Relevanz**? – Geldgeber:innen?

Welche **Kategorien** sind relevant?

Wie sind diese Kategorien definiert? (= operationalisiert)

Von wem werden diese Kategorien definiert?

Was passiert mit den Ergebnissen? Wem nützen die **Ergebnisse**?

→ Stichprobenauswahl?

→ Fallzahl?

→ Maßzahl?

→ Operationalisierung: was heißt: „*Arm*“ und „*reich*“ ?

Operationalisierung = Messbarmachen von Begriffen

Begriffe und **Messdimensionen** müssen exakt definiert sein.

Realdefinition: wahr oder falsch:

Erstsprache / Teilzeit oder Vollzeit / Bildungsabschluss

Nominaldefinition: weder wahr noch falsch: sondern:

eine Vereinbarung über den Zweck eines Begriffs oder einer Messdimension

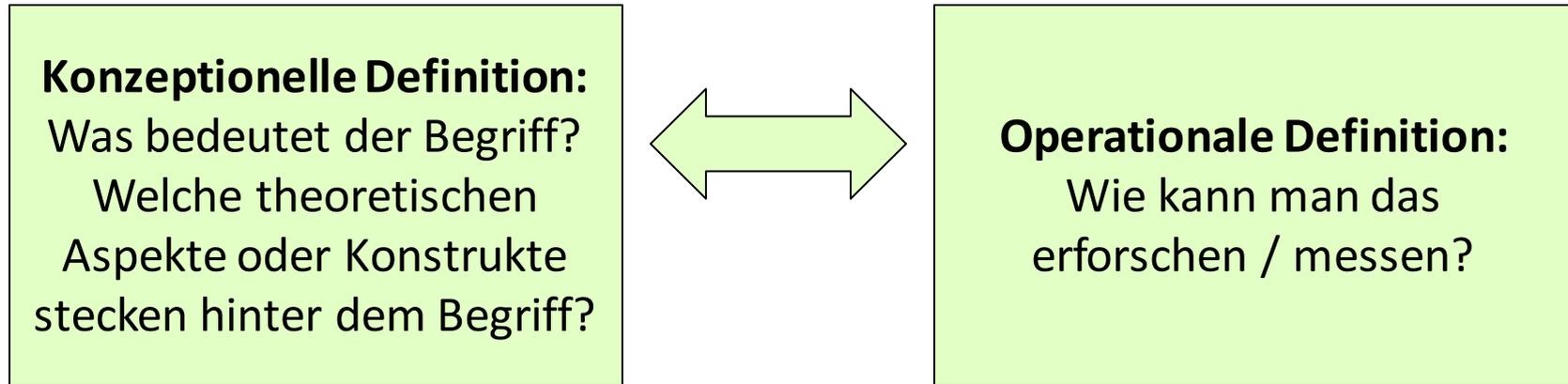
Kriterium ist die **Zweckmäßigkeit der Definition**

Begriff – z.B. „**Erstsprache**“: *ist die Sprache, die ein Kind als erste und wichtigste Sprache erlernt und verwendet (hat).*

z.B. „**Zweitsprache**“: *ist die Sprache, die eine Person nach der Erstsprache erlernt.*

z.B. „**starke Sprache**“: *ist die Sprache, die eine Person am besten spricht und am meisten nutzt.*

Begriffe / Konzepte in Forschungsfragen müssen zwei Prozesse durchlaufen:



Operationalisierung = das Messbarmachen von Begriffen

Begriffe sind meist umfassend und abstrakt und müssen in ihre Merkmale zerlegt werden, um sie messbar zu machen.

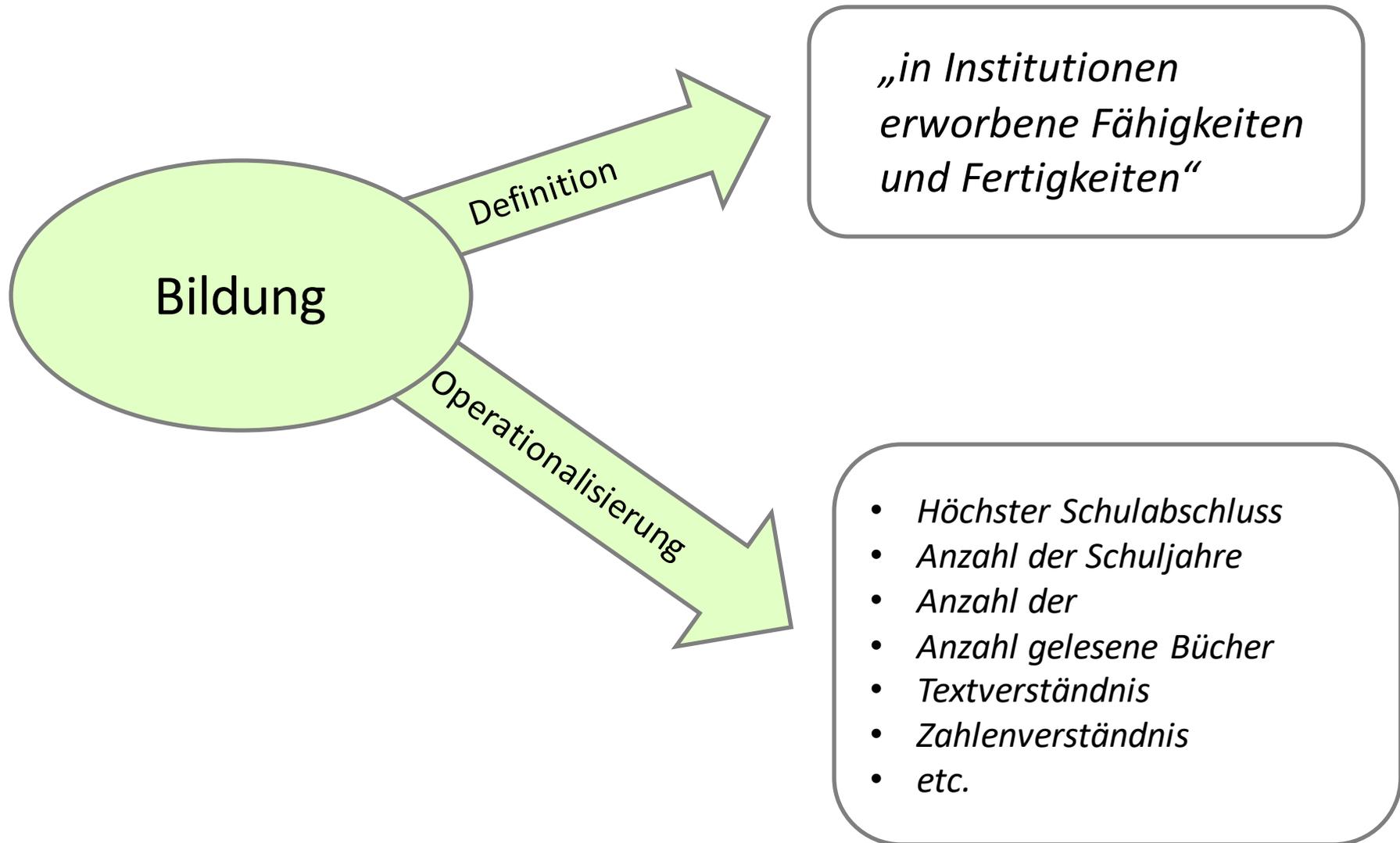
Begriffe und Messdimensionen müssen exakt definiert sein!

Voraussetzung ist die ausführliche Bedeutungsanalyse

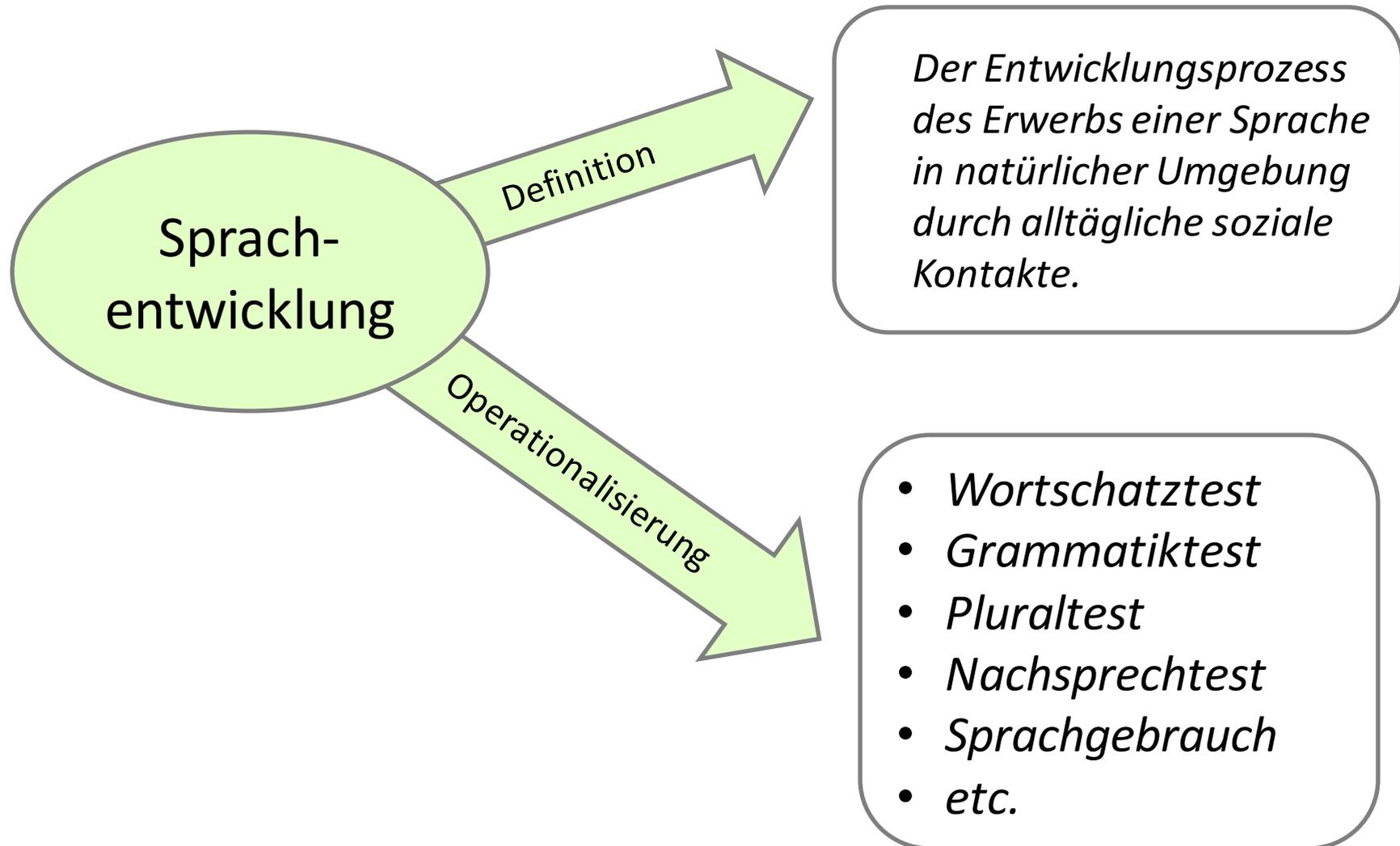
→ verschiedene Operationalisierungen sind möglich!

z.B. sprachliche Ausdrucksfähigkeit, Lebensqualität, Mehrsprachigkeit,...

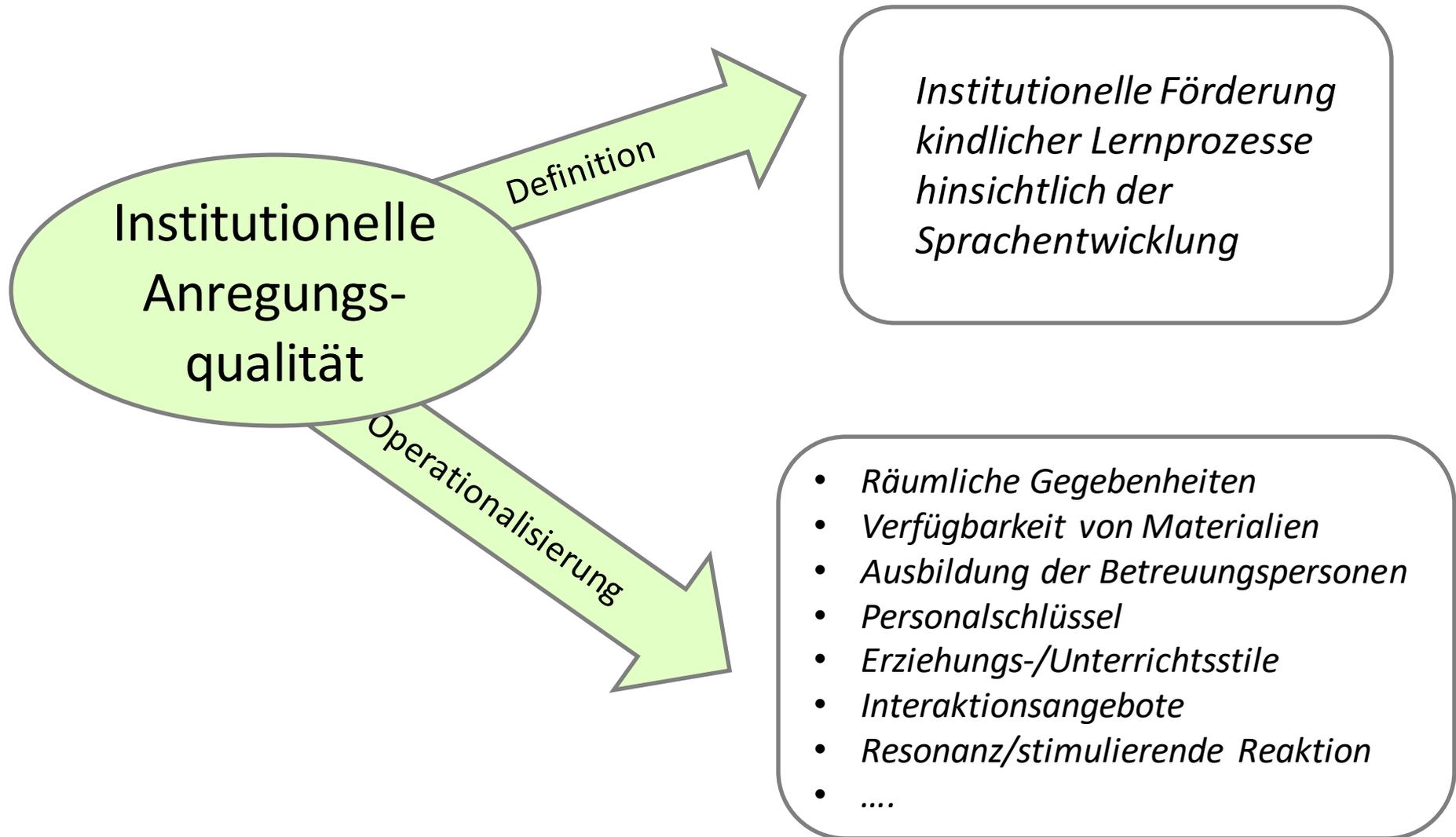
Beispiel Unterschied von Definition versus Operationalisierung



Beispiel Unterschied von Definition versus Operationalisierung



Beispiel Unterschied von Definition versus Operationalisierung



Kriterien einer Dimension

Eine **Dimension** besteht aus mehreren Items, welche alle „das Gleiche“ messen d.h. unter eine Überschrift passen (z.B. „*Lebensqualität*“).

Wichtig: **Eindimensionalität**

= die erhobenen Aspekte sollen wechselseitig miteinander zusammenhängen

→ muss zuerst inhaltlich und dann empirisch ermittelt werden

Messdimension – z.B. = Repräsentativ-Befragung (Ö) über die Gesundheit

Messdimension Lebensqualität

ergibt sich aus dem Grad der Zustimmung zu folgenden Aussagen:

Haben Sie genug Energie für das tägliche Leben?

Können Sie Ihr Aussehen akzeptieren?

Haben Sie genug Geld, um Ihre Bedürfnisse erfüllen zu können?

Haben Sie Zugang zu den Informationen, die Sie für das tägliche Leben brauchen?

Haben Sie ausreichend Möglichkeiten zu Freizeitaktivitäten?

Antwortmöglichkeiten: von 1 (=überhaupt nicht) bis 5 (= völlig)

Messdimension – z.B. = Untersuchung über die Gesundheit von Langzeitarbeitslosen:
Messdimension „Selbstsicherheit“

Selbstsicherheit

ergibt sich aus dem Grad der Zustimmung zu folgenden Aussagen:

Ich mag mich, so wie ich bin

Ich habe Angst, dass andere schlecht über mich reden

Ich traue mir viel zu

Ich zweifle an meinen fachlichen Fähigkeiten

Ich finde meist eine Lösung für schwierige Probleme

Ich schäme mich wegen meines Aussehens oder meiner Figur

Schwierigkeiten sehe ich gelassen entgegen

Ich finde Mittel und Wege, mich gegen Widerstand durchzusetzen

Antwortmöglichkeiten: von 1 (=nie) bis 6 (= immer)

Messdimension – z.B. = Untersuchung über die Lebenssituation von Älteren:

Messdimension „Angst“

Angst ergibt sich aus dem Grad der Einschätzung der Angst in den Bereichen:

Kriminalität

Umweltzerstörung

schlechte Wirtschaftslage/Rezession

Einsamkeit, alleine sein

Zunahme extremer politischer Tendenzen (Links- Rechtsruck)

Krieg

hoher Ausländeranteil in Ö

Armut/soziale Notlage

Arbeitslosigkeit

nicht mehr gebraucht zu werden

eigene Krankheit oder Behinderung

mein Tod

Krankheit oder Behinderung von nahen Angehörigen/FreundInnen

Tod von nahen Angehörigen oder FreundInnen

Antwortmöglichkeiten von 1 (= keine Angst) bis 7 (= große Angst)

Vorraussetzungen für die Operationalisierung

- theoretische Basis für die Fragestellung
- präzise Formulierung der Fragestellung und des Forschungsziels
- explorative Vorstudien
- Befragung von ExpertInnen
- Literaturrecherche
 - wie haben andere ForscherInnen diesen Sachverhalt erhoben?
- präzise Definitionen der verwendeten Begriffe

Erst nach der Definition der Begriffe erfolgt die Operationalisierung!!

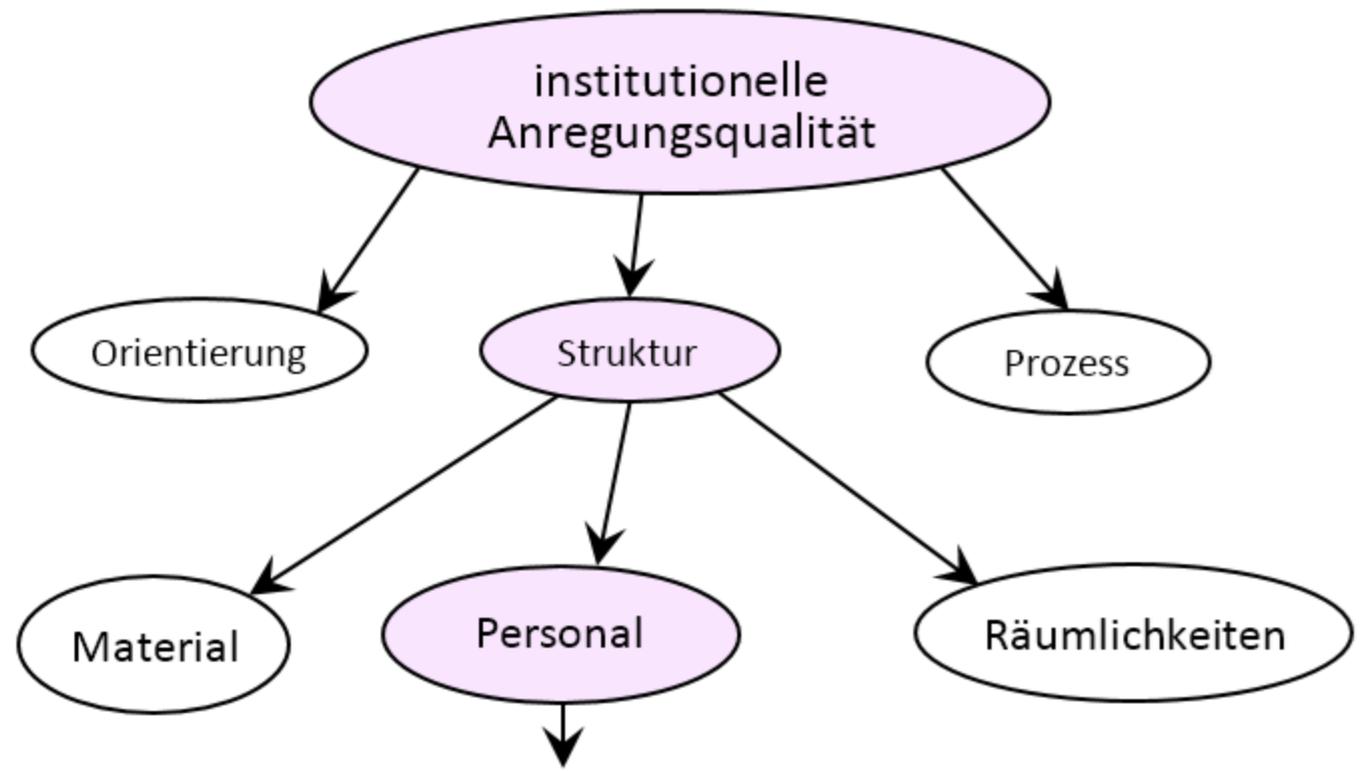
Konzept/Konstrukt – Dimension – Indikator

**Konzept
(=Konstrukt)**

**Dimensionen
(latent)**

**Sub-
Dimensionen
(latent)**

**Indikatoren
(manifeste
Merkmale)**



- Anzahl der Betreuungspersonen insgesamt
- Anzahl der Betreuungspersonen pro Gruppe
- Fluktuationsrate, Krankenstände,
- Ausbildungsgrad
- Arbeitszeitgestaltung
-

**Die Konstrukte in einen Zusammenhang bringen:
 Beispiel: Spracherwerb bei Kindergartenkindern
 Input – Prozess - Output**

INPUT	Prozess	OUTPUT
<p align="center">Erklärende – nicht beeinflussbar („Background“)</p>	<p align="center">Erklärende – beeinflussbar („Hebelpunkte“)</p>	<p align="center">Abhängige Faktoren („Ergebnis“)</p>
<p>herrschende Werthaltungen finanzielle Situation Erziehungsstil der Eltern Selbstwirksamkeit der Eltern </p>	<p>Nutzung der vorhandenen Materialien Anzahl der Interaktionen Verhalten des Umfelds </p>	<p>Wortschatz Sprachverständnis Expressiver Sprachgebrauch </p>

Was ist eine Hypothese?

„Wenn..... → dann.....“

und nicht: „Wie ist.....?“

Hypothese = Vermutung über einen Sachverhalt – muss überprüfbar sein!
Die Überprüfbarkeit ist der Unterschied zu „Alltagshypothesen“:

Alltagshypothesen

„Sind Beipacktexte für Medikamente verständlich?“

„Die meisten MitarbeiterInnen sind an Weiterbildungen desinteressiert.“

„Leuten, die lange arbeitslos sind, geht es schlecht.“

Wissenschaftliche Hypothesen

„Die Verständlichkeit von Beipacktexten ist abhängig von Gestaltung und Formulierung der Texte.“

„Je höher die Qualifikation der MitarbeiterInnen, desto häufiger besuchen diese Weiterbildungsveranstaltungen.“

„Je länger die Dauer der Arbeitslosigkeit, desto negativer die Erfolgserwartung, einen Arbeitsplatz zu bekommen.“

Was ist eine Hypothese?

Eine Hypothese ist eine Aussage über einen Sachverhalt, die überprüfbar ist.

Unterschied: Jugendliche aus der **Versuchsgruppe** (Therapie-Gruppe) weisen eine höhere *Konfliktfähigkeit* auf als Jugendliche aus der **Kontrollgruppe**.

Zusammenhang: Je älter die Jugendlichen, **desto** höher ist deren *Konfliktfähigkeit*.

Veränderung: Nach der Teilnahme an der Therapie-Gruppe hat sich die *Konfliktfähigkeit* der Jugendlichen **verbessert**.

„Wenn..... → dann.....“ und nicht: „Wie ist.....?“

Was ist ein Hypothesenmodell?

Eine Fragestellung ist die Zusammenstellung von einem sinnvollen Hypothesen-Set: Hypothesen, die inhaltlich in Zusammenhang stehen und aufeinander abgestimmt sind.

Erstellen eines Modells zur Hypothesenprüfung

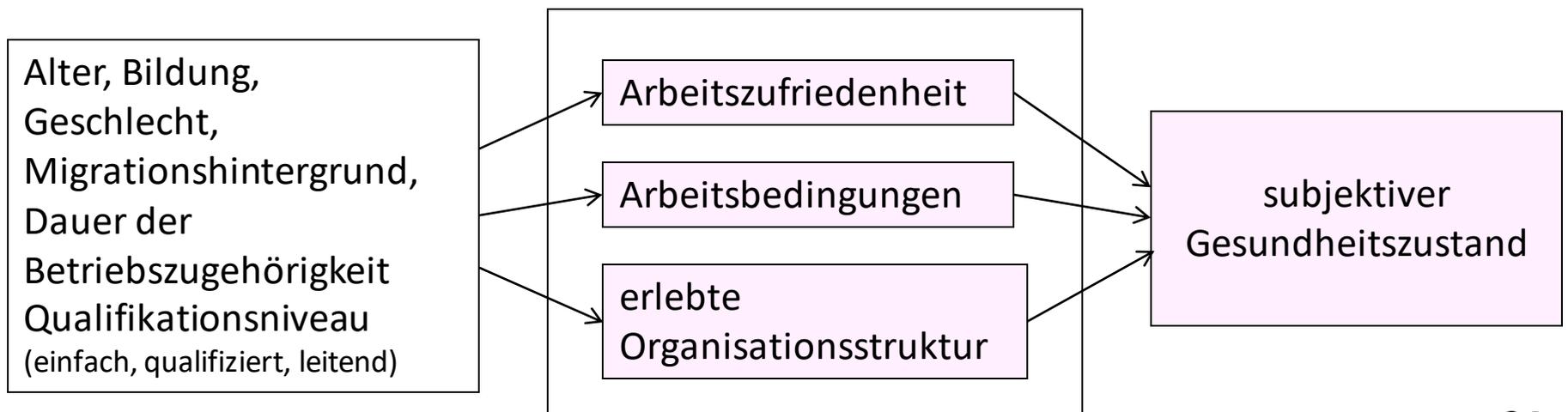
Hypothese =

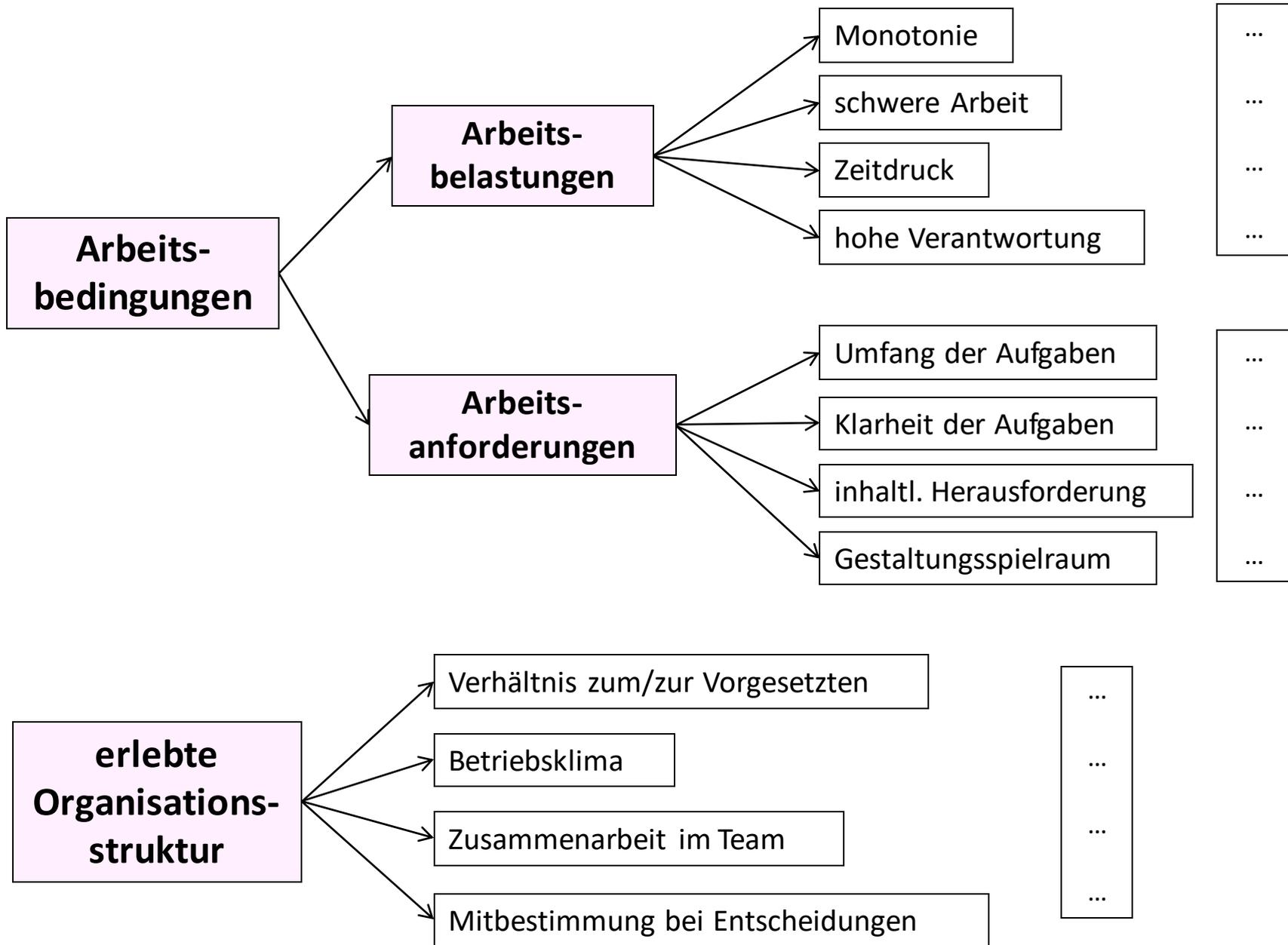
Vermutung über einen Gruppenunterschied / Zusammenhang zweier Merkmale

Fragestellung = mehrere Hypothesen in einem sinnvollen Zusammenhang

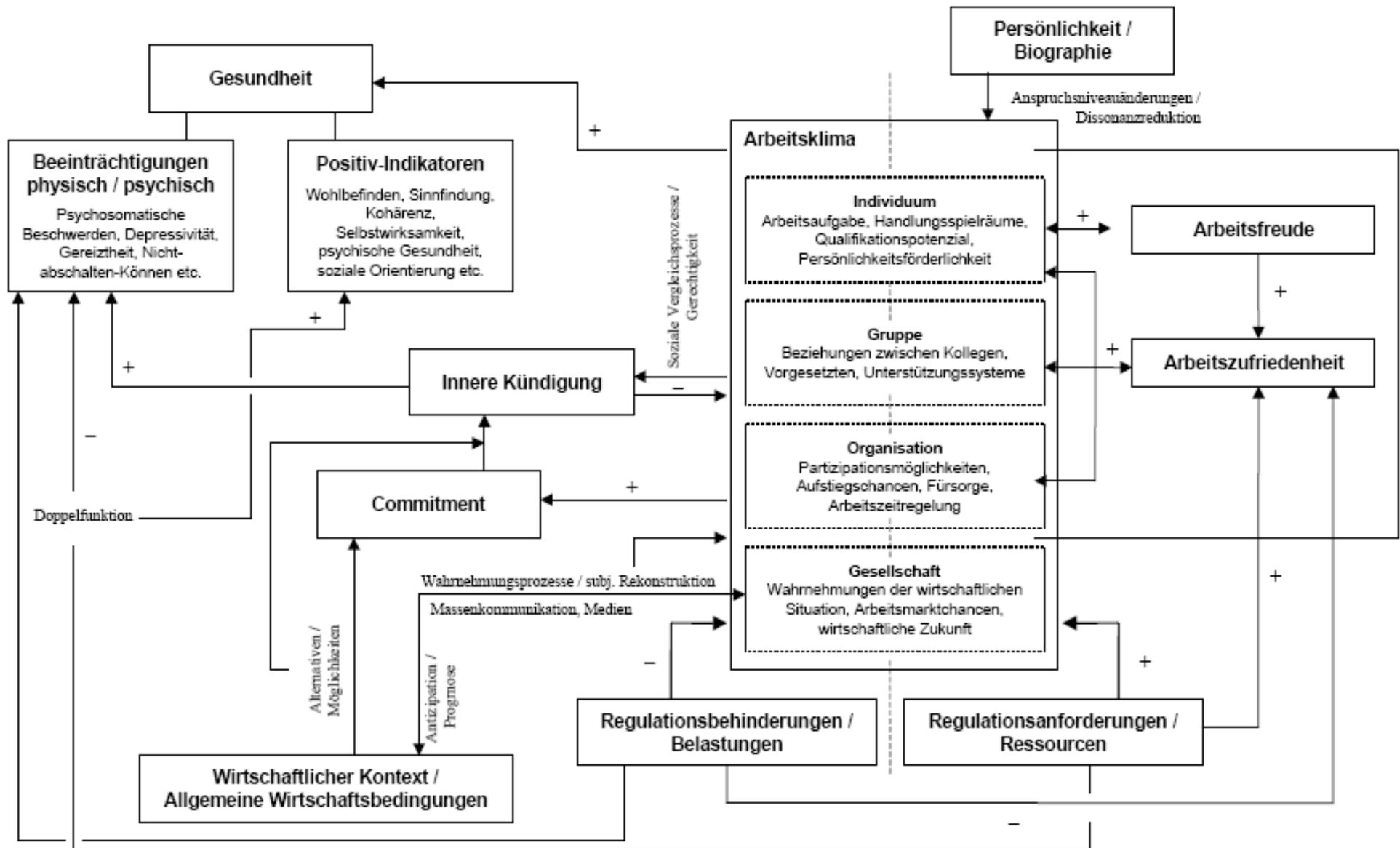
z.B. Ein **Papierproduktionsbetrieb** will Gesundheit und Arbeitszufriedenheit seiner Mitarbeiter*innen erheben:

Fragestellung: Beeinflussen Arbeitsbelastung, Arbeitszufriedenheit und Organisationsstruktur den Gesundheitszustand der Mitarbeiter*innen?

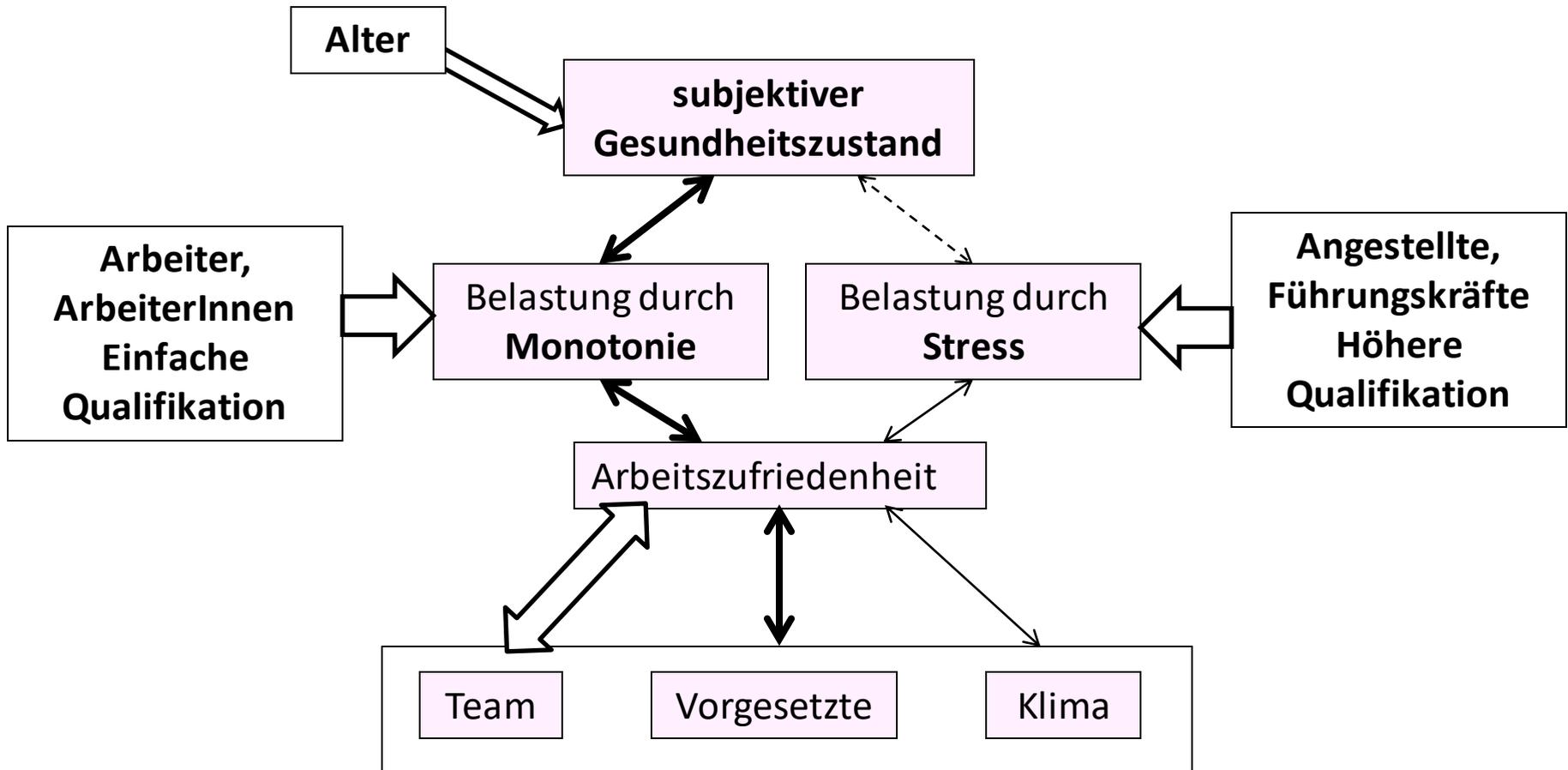




Arbeitsklima-Index - Rahmenmodell



Hypothesenmodell nach der Datenanalyse



Mögliche Hypothesen: Prinzip „*Wenn..., dann...*“

Frauen werden häufiger in einfachem Qualifikationsniveau eingesetzt als Männer.
MA mit **Migrationshintergrund** sind häufiger ArbeiterInnen als Angestellte.

oder Prinzip: „*Je..., desto...*“

Je einfacher die Tätigkeit, **desto** stärker sind die MA durch **Monotonie**, wenig inhaltlicher Herausforderung und wenig Gestaltungsspielraum belastet.

Je qualifizierter die Tätigkeit, **desto** stärker sind die MA durch **Stress**, hohe Verantwortung und Zeitdruck belastet.

Je älter die MitarbeiterInnen, **desto** schlechter bewerten Sie ihren Gesundheitszustand.

Je schlechter der Gesundheitszustand bewertet wird, **desto** stärker empfinden die MA die Belastung durch Monotonie und schwere körperliche Arbeit.

Je stärker die Belastung durch Monotonie, **desto** niedriger die Arbeitszufriedenheit.

Je besser die Zusammenarbeit im Team erlebt wird, **desto** höher ist die Arbeitszufriedenheit.

Definition von Begriffen

z.B. Nominaldefinition von
„**Armut**“ im SEB

Wie oft müssen Sie auskommen ohne ...
... ausreichende Mahlzeiten am Tag
... ausreichende Heizung
... ausreichende Kleidung

Antwortmöglichkeiten:
nie / kaum / manchmal / oft

Nominaldefinition von
„**Kriegsbetroffenheit**“ im SEB

Familienmitglied ...
... musste das Zuhause verlassen
... wurde verletzt
... wurde getötet

Antwortmöglichkeiten:
ja / nein

Operationalisierung bestimmt die Qualität des Messinstruments

Validität (Gültigkeit) = Instrument misst, was es messen soll
bzw. was es vorgibt, zu messen

Konstruktvalidität = bestmögliche Operationalisierung des „Konstrukts“

Vollständigkeit: Sind alle latenten Variablen enthalten? (Eigenschaften des Konstrukts)

Präzise: Ist das Konstrukt präzise und nachvollziehbar definiert und abgegrenzt?

Bsp. „*Armut*“ = Essen, Heizung, Kleidung (Was fehlt?)

„*Kriegsbetroffenheit*“ = Flucht, Einrücken, Verletzung, Tod (Was fehlt?)

Kriteriumsvalidität = hängt das Konstrukt mit anderen empirischen Kriterien zusammen?

Bsp. Äußere Krit-Val: „*Armut*“ = BIP, Arbeitslosenquote, Durchschnittseinkommen,
Konsumhäufigkeiten

Bsp. Innere Krit-Val: „*Armut*“ = Hängt der erhobene Armuts-Index mit dem erhobenen
HH-Einkommen zusammen? (z.B. mit dem Freizeitverhalten)

Operationalisierung bestimmt die Qualität des Messinstruments

Konvergenzvalidität

= Messdimensionen, die dasselbe Konstrukt messen, müssen hoch miteinander korrelieren

Diskriminanzvalidität

= Messdimensionen, die verschiedene Konstrukte messen, müssen gering/nicht miteinander korrelieren

Reliabilität (Zuverlässigkeit) = Instrument erbringt bei wiederholtem Einsatz gleiche Ergebnisse (Standardisierte Skalen)

Bsp. „*Armut*“ = Ergebnis abhängig von Jahreszeit ??
Reliabilität ist durch statistische Verfahren beurteilbar.

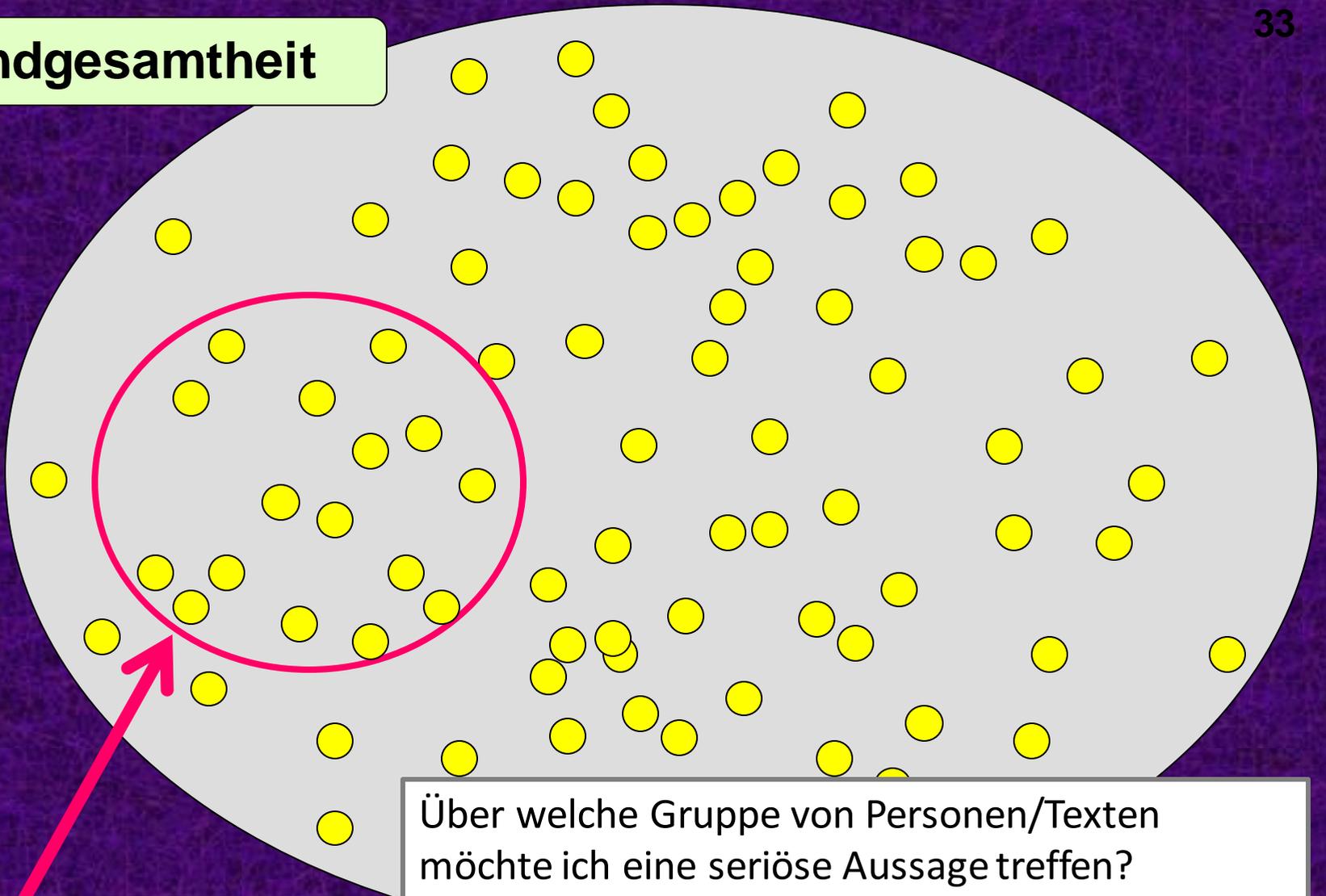
Reliabilitätsanalyse

= Messdimensionen, die dasselbe Konstrukt messen, müssen eindimensional sein, d.h. sie dürfen keine Subdimensionen enthalten.

Grundgesamtheit und Stichprobe

Stichprobendesign

Grundgesamtheit



Stichprobe

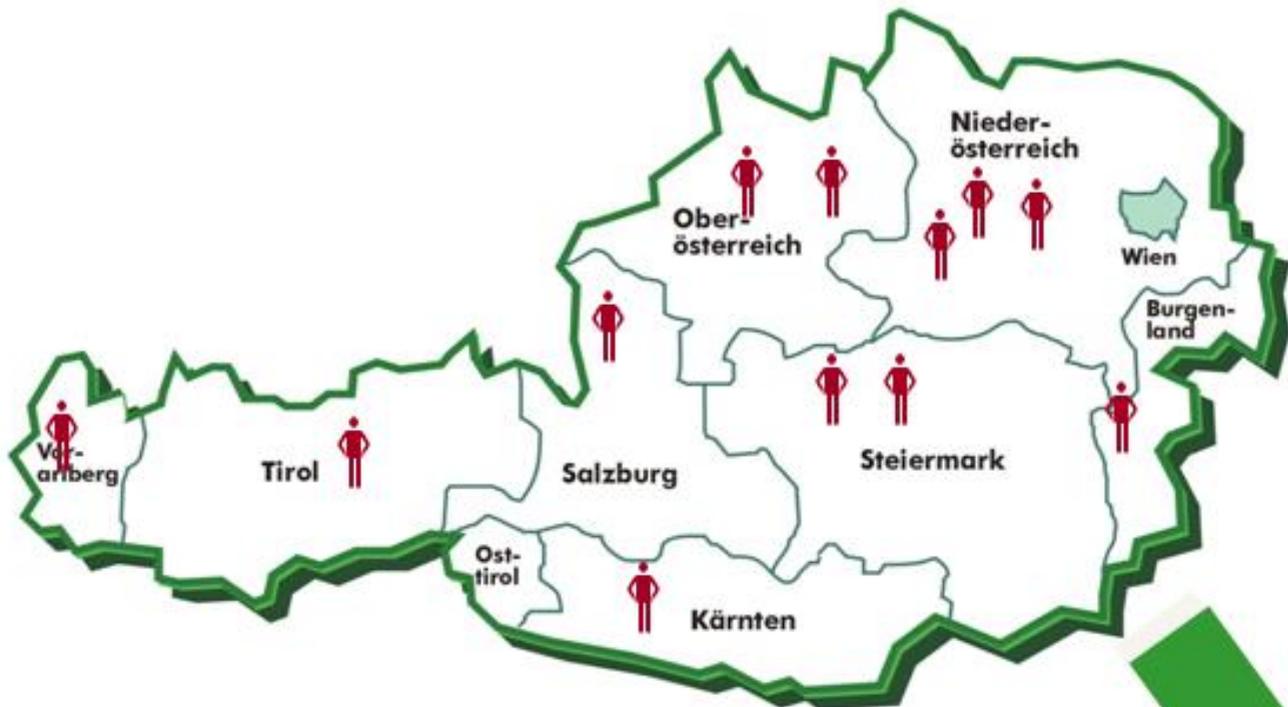
Über welche Gruppe von Personen/Texten möchte ich eine seriöse Aussage treffen?
ODER: Für welche Gruppe von Personen/Texten sollen meine Ergebnisse gelten?
Wen soll ich befragen/untersuchen?

UNIVERS = Grundgesamtheit (=Population)

Menge aller Elemente (Personen, Texte, Gespräche),
für die eine Fragestellung beantwortet werden soll
= für die die Ergebnisse gelten sollen.

Eine exakte Eingrenzung zu Beginn eines Forschungsprojektes
ist sehr wichtig!

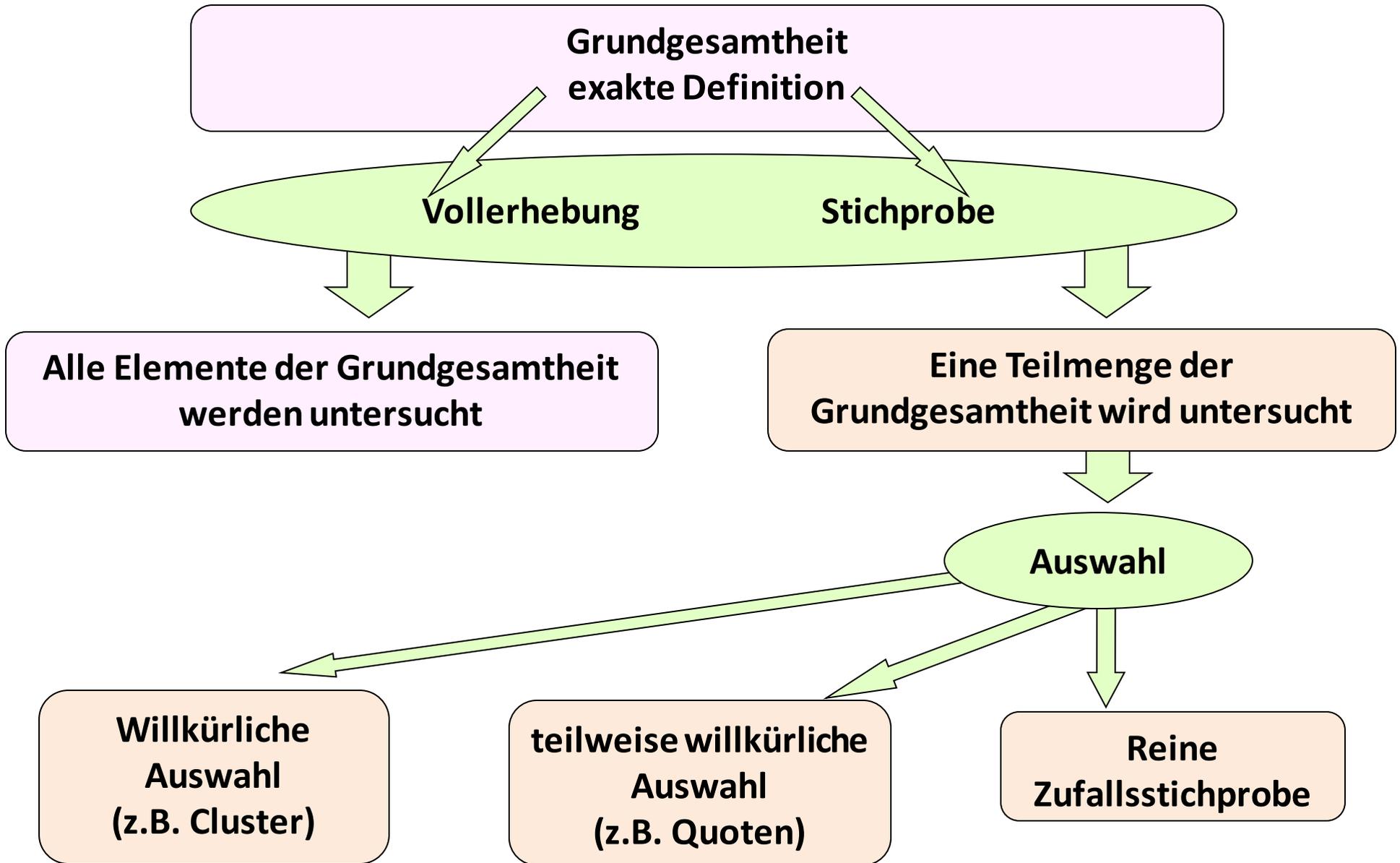
- z.B.: Alle Studierenden der Sprachwissenschaft im Jahr xxxx
- z.B.: Alle TeilnehmerInnen an der VO Einführung in die Statistik
im SoSe xxxx
- z.B.: Alle Personen ab 16 Jahren, die in Österreich leben
und zweisprachig aufgewachsen sind, im Jahr xxxx



UNIVERS
6.525.400 Personen 15+

Sample
1000 Interviews 15+





Grundgesamtheit (Population) versus Stichprobe

Grundgesamtheit = Die Menge aller Elemente/Personen, für die die aufgestellte Fragestellung beantwortet werden soll

z.B. österreichische Bevölkerung ab 16 Jahre im Jahr 2018

Wichtig: genaue Definition – exakte Eingrenzung zu Beginn der Forschungsarbeit

Stichprobe = Auswahl aus der GG; **Repräsentativität** = Miniaturabbild der GG

Stichprobe entspricht in allen relevanten Merkmalen der GG;

Auswahl erfolgt nach bestimmten Methoden: Stichprobenziehung

Stichprobenziehung = Vollerhebung: nicht immer möglich bzw. sehr teuer

→ **Reine Zufallsstichprobe** (Random sample)

→ **Quotenstichprobe** (aus vorgegebenen Quoten Zufallsstichproben ziehen)

→ **Clusterstichprobe** (alle Elemente aus mehreren klar umschriebenen Einheiten)

Systematische Fehler = bestimmte Gruppen sind über- oder unterrepräsentiert

= Bias = „Verzerrung“ (der Ergebnisse)

z.B. bei Telefoninterviews: mehr Frauen, Ältere, Ärmere

z.B. bei Evaluierung von KundInnenzufriedenheit: Unzufriedene: weniger Rücklauf

Zufallsstichprobe

Reine Zufallsstichprobe:

Optimalfall (selten), nur möglich, wenn **alle Elemente aus der GG bekannt** sind.

meist: Telefonische Interviews mit Zufallsgenerator

Bester Garant von **Repräsentativität** ist der Zufall

Eigenschaften von Zufallsstichproben:

- jedes Element der GG besitzt die gleiche Wahrscheinlichkeit, in die Stichprobe zu kommen („gezogen zu werden“)
- jede mögliche Stichprobe besitzt die gleiche Wahrscheinlichkeit, gezogen zu werden (d.h. jede mögliche Stichprobe bringt gleiche Ergebnisse)
- dadurch wird es möglich, **von der Stichprobe auf die GG zu schließen** und Zufallsfehler (Schätzfehler, Schwankungsbreiten, Standardfehler) zu berechnen

**Rückschlüsse von der Stichprobe auf die GG sind nur dann zulässig,
wenn die Stichprobe repräsentativ ist**

(= ein Miniaturabbild der GG ohne Verzerrungen)

Beachte: 100%ige Repräsentativität ist nicht erreichbar!

Grundgesamtheit (Population) versus Stichprobe

Wie groß soll/muss die Stichprobe sein?

Die Größe der Stichprobe bestimmt sich aus der erwünschten Genauigkeit der Ergebnisse (tolerierte Schwankungsbreiten) – wie viel Schwankungsbreite will ich tolerieren?

Das heißt: Je größer die Stichprobe, desto weniger zufallsabhängig sind die Ergebnisse.

Grundsätzlich gilt: Je mehr – desto besser! Orientiert sich meist am vorhandenen Budget.

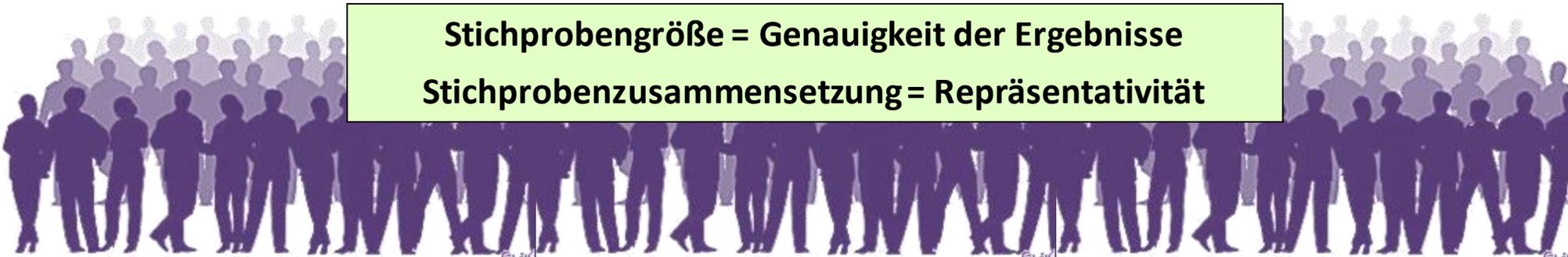
Faustregel: Für jede Gruppe, die ich untersuchen will, MINDESTENS 30 Personen einplanen (besser: 50 Personen, weil: oft fehlende Angaben).

Z.B. Ich vergleiche 3 Altersgruppen nach Geschlecht = $3 \times 2 \times 30 = 180$ P.

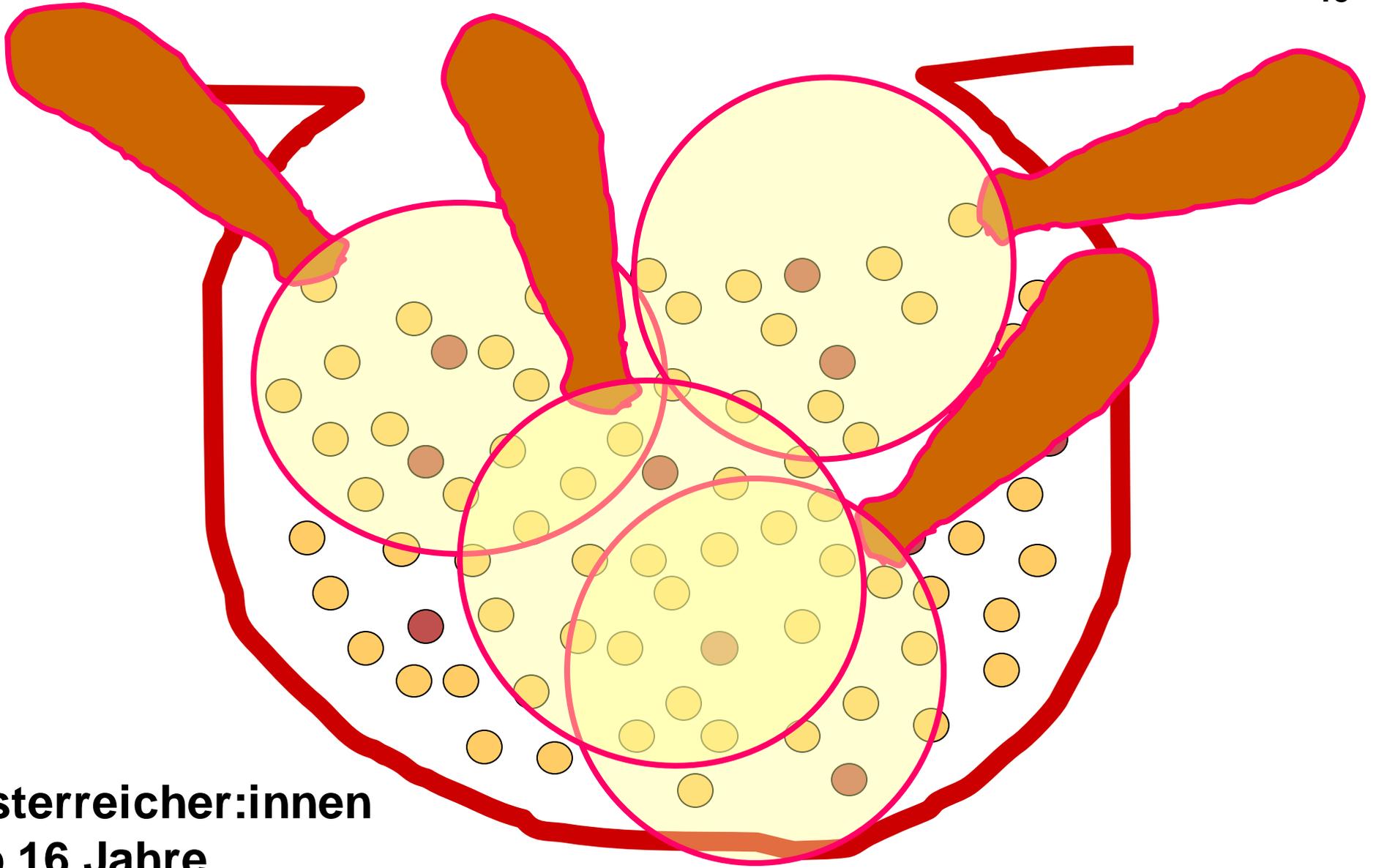
(„Zentraler Grenzwertsatz“ = Je größer die Fallzahl, desto kleiner die Zufallsfehler.)

BEACHTEN: Die Größe der Stichprobe hat NICHTS zu tun mit der Repräsentativität!

→ Nur weil eine Stichprobe groß ist, bedeutet das nicht, dass sie auch repräsentativ ist!



Stichprobengröße = Genauigkeit der Ergebnisse
Stichprobenzusammensetzung = Repräsentativität



**Österreicher:innen
ab 16 Jahre
im Jahr xxxx**

27% der Österreicher:innen sind zweisprachig aufgewachsen.

Schwankungsbreiten (KI mit 95%-iger Sicherheit)

	%	%	%	%	%	%	%	%	%	%
	3	5	10	15	20	25	30	40	50	
n	97	95	90	85	80	75	70	60	50	
100	3,4	4,4	6,0	7,1	8,0	8,7	9,2	9,8	10,0	
200	2,4	3,1	4,2	5,0	5,7	6,1	6,5	6,9	7,1	
300	2,0	2,5	3,5	4,1	4,6	5,0	5,3	5,7	5,8	
400	1,7	2,2	3,0	3,6	4,0	4,3	4,6	4,9	5,0	
500	1,5	1,9	2,7	3,2	3,6	3,9	4,1	4,4	4,5	
750	1,2	1,6	2,2	2,6	2,9	3,2	3,3	3,6	3,7	
1.000	1,1	0,4	1,9	2,3	2,5	2,7	2,9	3,1	3,2	
1.250	1,0	1,2	1,7	2,0	2,3	2,4	2,6	2,8	2,8	
1.500	0,9	1,1	1,5	1,8	2,1	2,2	2,4	2,5	2,6	
2.000	0,8	1,0	1,3	1,6	1,8	1,9	2,0	2,2	2,2	
2.500	0,7	0,9	1,2	1,4	1,6	1,7	1,8	2,0	2,0	
3.000	0,6	0,8	1,1	1,3	1,5	1,6	1,7	1,8	1,8	
3.500	0,6	0,7	1,0	1,2	1,4	1,5	1,5	1,6	1,7	
4.000	0,5	0,7	0,9	1,1	1,3	1,4	1,4	1,5	1,6	
5.000	0,5	0,6	0,8	1,0	1,1	1,2	1,3	1,4	1,4	
7.500	0,4	0,5	0,7	0,8	0,9	1,0	1,1	1,1	1,2	
10.000	0,3	0,4	0,6	0,7	0,8	0,9	0,9	1,0	1,0	

Wie viel Schwankungsbreite möchte ich tolerieren?

Lesebeispiel zum Konfidenzintervall

(=Schwankungsbreiten, = Zufallsfehler, mit dem zu rechnen ist)

27 von 100 Befragten (= 27%) sind zweisprachig aufgewachsen.

Zeile: $n = 100$, Spalte: 30:70 → Fehler $\pm 9,2\%$

Wahrer Anteil in der GG schwankt zwischen **17,8% bis 36,2%**.

405 von 1500 Befragten (= 27%) sind zweisprachig aufgewachsen.

Zeile: $n = 1500$, Spalte: 30:70 → Fehler $\pm 2,4\%$

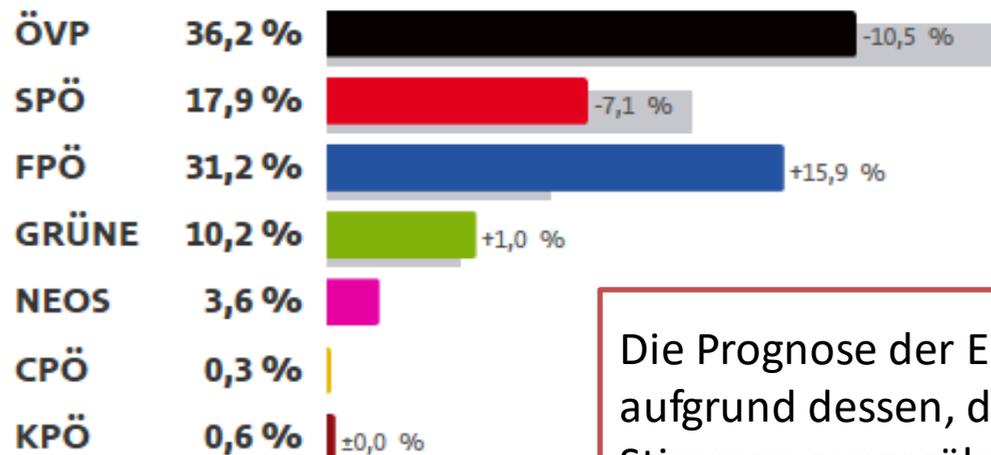
Wahrer Anteil in der GG schwankt zwischen **24,6% bis 29,4%**.

= Zwischen 24,6% und 29,4% der Österreicher:innen sind zweisprachig aufgewachsen.

Wie viel Schwankungsbreite möchte ich tolerieren?

Hochrechnung Landtagswahl von 16.40 Uhr

Stimmenanteile



Schwankungsbreite: **±2,6 %**
Ausählungsgrad: 22,5 %
Wahlbeteiligung: 82,3 %

Die Prognose der Ergebnisse beinhaltet aufgrund dessen, dass noch nicht alle Stimmen ausgezählt wurden, noch eine relativ große Schwankungsbreite! Das Ergebnis der ÖVP z.B. könnte theoretisch noch zwischen 33,6% und 38,8% liegen.

Reine Zufallsstichprobe



- Liste aller Personen der Grundgesamtheit
- Erstellung einer Liste von Zufallszahlen
- Auswahl
- Einfacher: systematische Auswahl
- z.B. jede 100. Person auf der Liste

Reine Zufallsstichprobe

Grund-
gesamtheit

Liste aller einzelnen Elemente (z.B. Personen) der Grundgesamtheit



Stichprobe

Ziehung einer Zufallsstichprobe aus allen Elementen der Grundgesamtheit





Geschichtete Zufallsstichprobe

Grundgesamtheit liegt in Schichten vor.
 Schicht = eine Gruppe mit einem
 gemeinsamen Merkmal
 z.B. Erstsprache / Religionsbekenntnis (...)
 Aus jeder Schicht wird eine Zufallsstich-
 probe gezogen.

Geschichtete Stichprobe

Grund-
gesamttheit

Schicht 1

Schicht 2

Schicht 3

Schicht 4

Schicht 5



Stichprobe



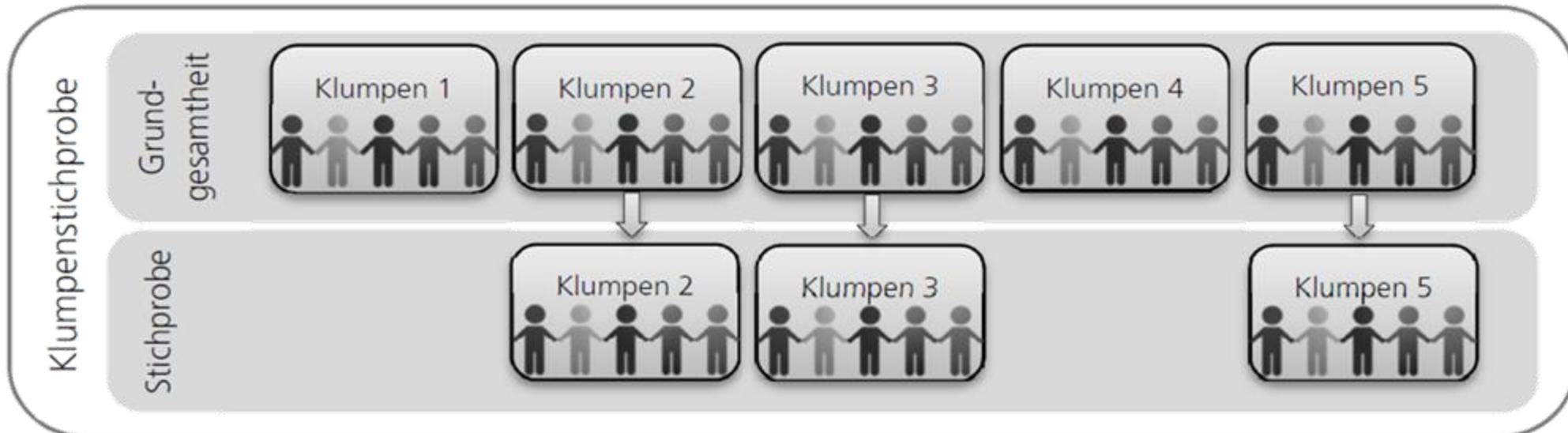
Clusterstichprobe/Klumpenstichprobe

Die GG liegt in Clustern vor, also Clustern von Personen z.B. eine Schule, ein Betrieb, ein Bundesland..

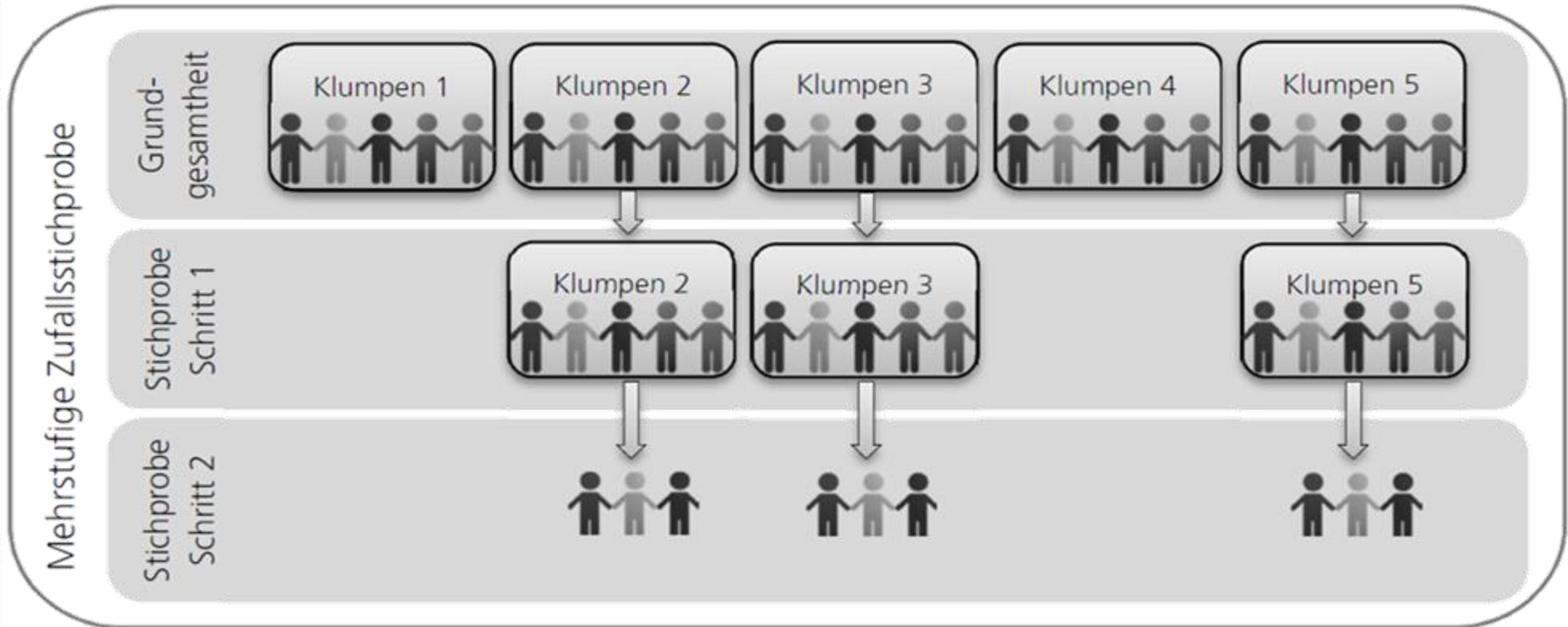
Aus den Clustern von Personen werden einzelne ausgewählt

wie z.B. Schulklassen, Abteilungen, Gemeinden..

Die Cluster werden total erhoben oder zufällige Auswahl von Personen.



Mehrstufige Clusterstichprobe



Die GG liegt in Clustern vor. (z.B. Schulen, Betriebe..)

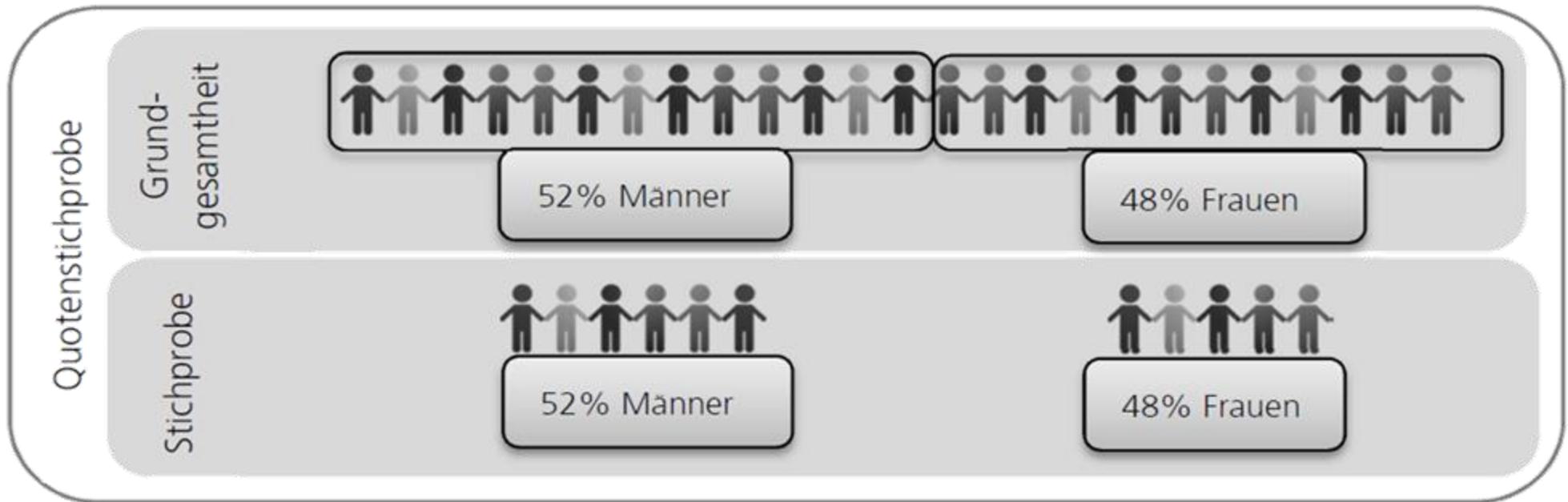
Aus den Clustern werden Cluster bewusst oder zufällig ausgewählt.

Aus den bestimmten Clustern werden wieder Subcluster bestimmt.

(z.B. Stationen, Schulklassen, Abteilungen...)

Die Subcluster werden total erhoben oder zufällige Auswahl von Personen.

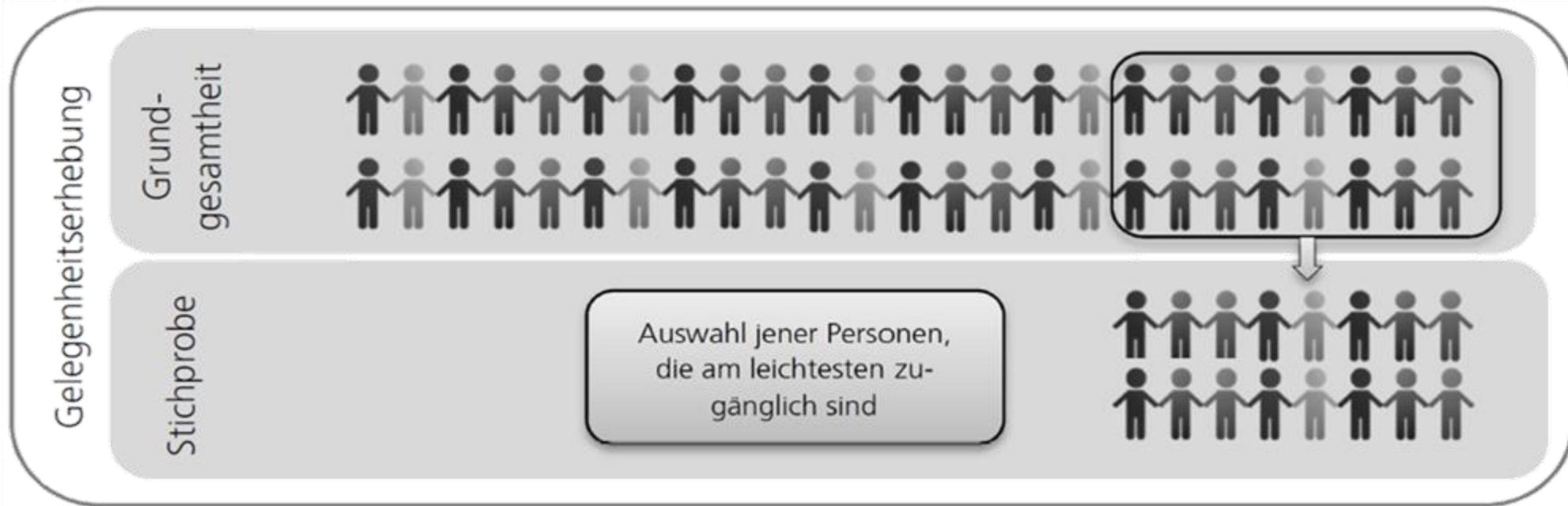
Quotenstichprobe



Die GG wird in Quoten aufgeteilt. Innerhalb dieser Quoten wird so lange nach dem Zufallsprinzip erhoben, bis die vorgegebene Quote erfüllt ist.
Typisch für Erhebungen durch Meinungsforschungsinstitute.
Ist keine reine Zufallsstichprobe, wird aber als solche behandelt.....

Gelegenheitsstichprobe

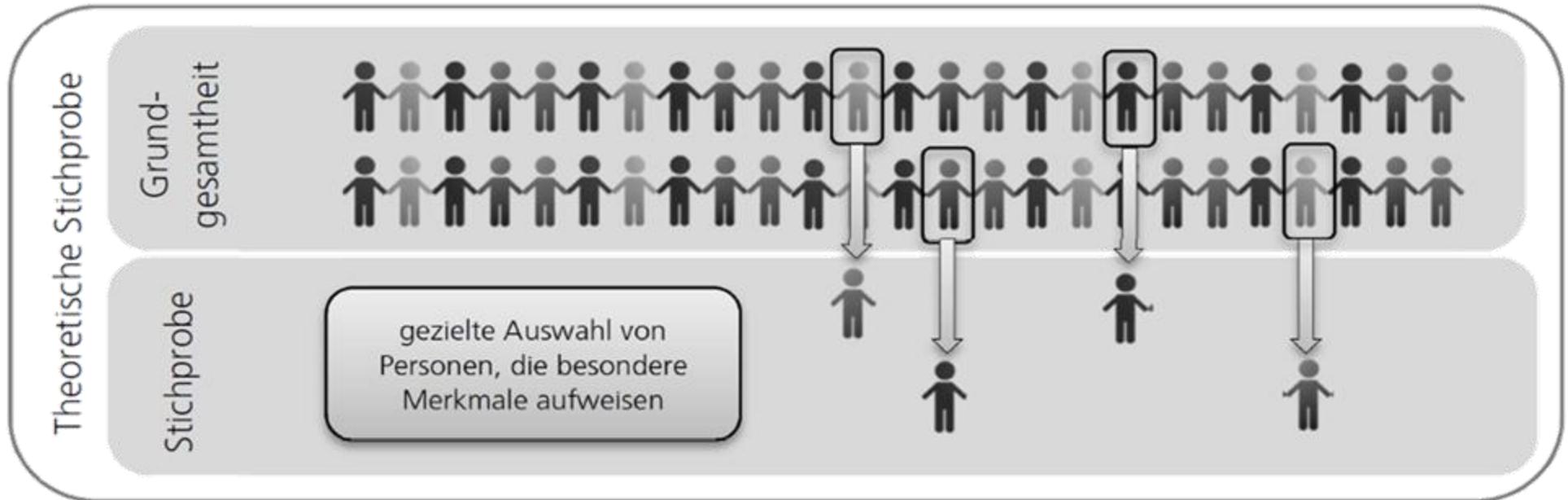
explorativ, keine Repräsentativität, kein Zufall!



Personen, die am leichtesten zugänglich sind werden ausgewählt
(z.B. die nächsten 100 Personen, die ich auf der U-Bahnstation XY antreffe)
Spezialfall: convenience sample: Befragung von Freund*innen, Bekannten,...

Theoretische Stichprobe

explorativ, keine Repräsentativität, kein Zufall!



Für eine bestimmte Population typische Personen oder Personen einer „ungewöhnlichen Gruppe“ werden ausgewählt.

Überblick Stichprobenarten

Stichprobe	GG	Stichprobe 1	Stichprobe 2	Stichprobe 3	Rückschluss auf GG
Einfache Zufallsstichprobe	Liste von Personen (z.B. KlientInnen von 2005-2015)	Zufallsauswahl von Personen ausgewählt (z.B. jede 10. Person)			✓
Geschichtete Zufallsstichprobe	Personengruppe in Schichten z.B. Personen in den Pflegestufen	Zufallsauswahl von Personen pro Schicht			✓
(Einfache) Clusterstichprobe	Mehrere Cluster (z.B. Klassen einer Schule)	Zufallsauswahl von Clustern	Totalerhebung der Personen in den Clustern		✓
Mehrstufige Clusterstichprobe	Mehrere Cluster (z.B. Schulen)	Zufallsauswahl von Clustern (manchmal auch willkürlich)	Zufallsauswahl von Subclustern ausgewählt (z.B. Klassen)	Total- oder Zufallserhebung der Personen in den Subclustern	✓
Quotenstichprobe	Personengruppe mit bestimmten Quoten	Zufallsauswahl von Personen, bis die Quoten erreicht sind, dann willkürlich			eigentlich nicht zulässig, wird aber gemacht
Gelegenheitsstichprobe	Personengruppe mit bestimmten Merkmalen (meist keine Liste möglich)	willkürliche Auswahl jener Personen, die am leichtesten zugänglich sind			⊘
Theoretische Stichprobe	Personengruppe mit bestimmten Merkmalen (meist keine Liste möglich)	willkürliche Auswahl von typischen oder atypischen "Fällen"			⊘

Überblick Stichprobenarten

Zufallsstichproben (Zufall, teilweise willkürlich)

Reine Zufallsstichprobe

Klumpenstichprobe

Mehrstufige
Klumpenstichprobe

Quotenstichprobe

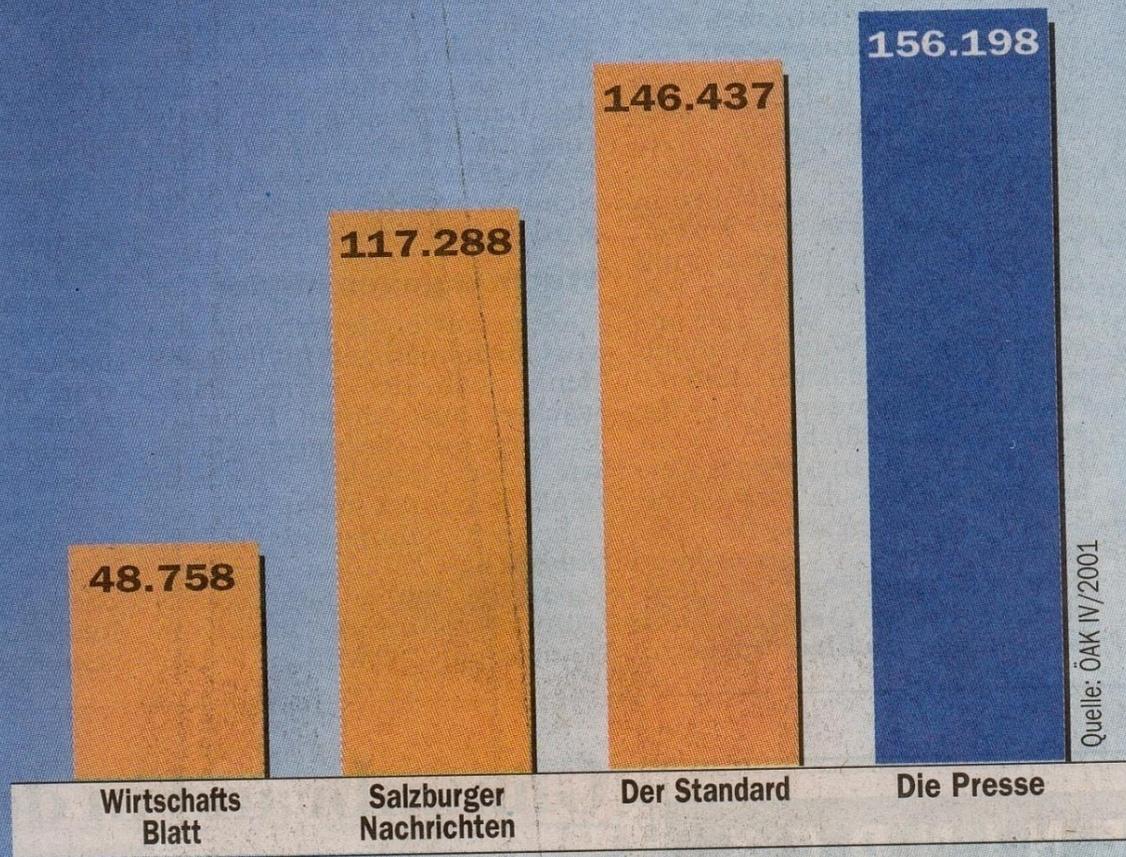
Bewusste Stichproben (willkürlich)

Gelegenheitserhebung

Theoretische Stichprobe

Die bewusste Erhebung (*Nicht-Zufallsstichprobe*)

- Verläuft nach weniger strengen Regeln
- Weniger repräsentativ – Streng genommen ist kein Schluss auf die Grundgesamtheit möglich.
- In der Praxis (auch im Bereich der sprachwissenschaftlichen Forschung) sind aber Zufallsstichproben oft nicht möglich.
- Repräsentativität ist außerdem nicht immer ausschlaggebend (z.B. bei explorativen Studien)



Verbreitete Auflage Inland, nur am Samstag

Quelle: ÖAK IV/2001

Datenquelle =
 ÖAK IV 2001
 „*Verbreitete Auflage am Samstag*“

Manipulative Darstellung von Ergebnissen

März 2001

So viel Zeitung muß sein.

Die Presse

DER STANDARD mit 401.000 Lesern

Regelmäßige Leser



Quelle: Media Analyse 2001

DER STANDARD

Datenquelle =

Media-Analyse 2001

IFES

*„regelmäßige
Leser“*

Manipulative Darstellung
von Ergebnissen

März 2001

LeserInnen pro Ausgabe Wien 14+



HEUTE
Gratiszeitung

*„LeserInnen pro Ausgabe
in Wien über 14 J.“*

Manipulative Darstellung
von Ergebnissen

Okt 2012

LeserInnen pro Ausgabe national 14+



Manipulative Darstellung von Ergebnissen Okt 2012

Dienstboten	1694	oder	25,5	Prozent
Gewöhnliche Arbeiter	3651	„	55,0	„
Handwerker und Fabrikarbeiter	974	„	14,7	„
Arbeitsscheue	320	„	4,8	„

USA 1890
 (aus: Der Verbrecher, Band II,
 Cesare Lombroso 1890)

Berufslose, Bettler, Vagabunden, Prostituierte, Spitalinsassen	59,2
Dienstboten	25,9
Landwirtschaftliche Bevölkerung	24,3
Gewerbe- u. handeltreibende Bevölkerung	18,1
Höhere Berufsarten	10,6

Frankreich 1880
 (aus: Der Verbrecher, Band II,
 Cesare Lombroso 1890)

„Aufenthaltsstatus“:
 Arbeitnehmer, Studenten, Selbständige
 Familiengemeinschaft mit Österreicher
 Touristen
 Asylwerber, Fremde ohne Beschäftigung,
 Nicht rechtmäßiger Aufenthalt

Österreich seit 2000
 (aus: PKS = polizeiliche
 Kriminal Statistik)

„Umfrage“

Stichprobe und Studiendesign

Repräsentativ (*deskriptiv*) versus experimentell (*kausal*)

Stichprobe repräsentiert GG

(in allen relevanten Merkmalen – Welche Merkmale sind relevant?)

Üblicherweise: Geschlecht, Alter, Bildung, Einkommen, Region, Erwerbstätigkeit

Möglichst viele – Begründung, warum ein Merkmal nicht relevant ist

Ergebnisse sind auf GG generalisierbar

Bedingung: **Zufallsauswahl** (Liste, Urne... – alle haben gleiche Chance)

Heterogene Stichprobe (möglichst viele Eigenschaften wie in GG – Rückschluss auf größere Personengruppe möglich.

Repräsentativität als Kontinuum

Je größer und heterogener die Stichprobe, je „zufälliger“ die Auswahl, desto „höher“ die Repräsentativität.....

Stichprobe und Studiendesign

„Vergleichsgruppen“

Repräsentativ (*deskriptiv*) versus **experimentell (*kausal*)**

Stichprobenauswahl durch Variation des zu untersuchenden Kriteriums

Anhand der Varianz dieses Kriteriums wird Kausalität unterstellt.
„Bedingungen“ – Medikament, Text, Film, Unterrichtsmethode.....

Gruppen unter verschiedenen Bedingungen:

Bedingung A – nicht A (Untersuchungsgruppe – Kontrollgruppe)
oder Bedingung A – B – C

Homogene Gruppen (möglichst viele Eigenschaften gleich)

– Einfluss der Bedingung soll „sauber“ bleiben, soll kontrolliert werden

Randomisierung: Gruppenteilung per Zufall (Studis auslosen)

Parallelisierung: homogene Gruppe (Germanistik-Studis 4. Sem.)

Matched pairs: „Gleiche Partner“ finden

Externe Validität – hohe Generalisierbarkeit (Feldexperiment)

Interne Validität – hohe Kontrolle (Laborexperiment)

– eindeutige, streng kontrollierte Bedingungen bei wenig Störvariablen

Eine „quantitative“ Fragestellung.....

... muss im Rahmen der Studie realistisch und seriös beantwortbar sein

Welchen Effekt hätte die Einführung eines Gratis-Mittagessen für alle Studierenden auf deren Lernerfolg während des gesamten Studiums?

In welchem Zusammenhang steht das Ernährungsverhalten von Studierenden mit deren Lernerfolg?

... muss mit quantitativen Methoden beantwortbar sein (messen und zählen!)

Welche Bedeutung hat für Studierende das Besuchen von Bibliotheken?

Gibt es Unterschiede in der Häufigkeit des Bibliotheksbesuchs von Studierenden nach soziodemografischen Merkmalen?

... muss als Ziel die Beschreibung und Erklärung eines Phänomens haben – und nicht das Verstehen!

Wie wirken sich verschiedene Studienverläufe auf den Lebenslauf von AbsolventInnen aus?

Wie stark beeinflussen die Rahmenbedingungen während des Studiums den Studienverlauf von Studierenden?

Eine „quantitative“ Fragestellung.....

.... muss begrenzt und präzise sein

Warum Studieren ältere Menschen?

In welchem Zusammenhang stehen Studienwahlmotive mit dem Alter bzw. der Biographie von Studierenden?

.... darf nur eine Frage beinhalten (notfalls die Fragestellung aufsplitten!)

In welchen Bereichen möchten Studierende der Sprachwissenschaft nach Ende des Studiums arbeiten, und in welchem Zusammenhang steht dies mit den gesetzten Schwerpunkten, dem Studienfortschritt und der durchschnittlichen Leistungsbeurteilung?

In welchem Zusammenhang stehen die gesetzten Schwerpunkte während des Studiums mit den subjektiv eingeschätzten Berufsaussichten?

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

Sprachgebrauch: Repräsentativbefragung von LehrerInnen und SchülerInnen
an österreichischen Schulen
LehrerInnen n = 164; SchülerInnen n = 1.264 (2015)

„Österreichisches“, „deutsches“ und „schweizerisches“ Deutsch:
Inwiefern unterscheiden sich das Sprachverständnis und der Sprachgebrauch
zwischen LehrerInnen und SchülerInnen?

Stehen Sprachverständnis und Sprachgebrauch in Zusammenhang mit sozialen
Merkmale (z.B. Geburtsort/Geburtsland)?

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

Spracherwerb: Kontrollierte Feldstudie mit mehreren Messzeitpunkten
n = 57 (2013 bis 2017)

Auswirkungen des sozialen Hintergrunds der Herkunftsfamilie auf die sprachlichen Kompetenzen (Grammatik und Wortschatz) bei Kindern im Kindergartenalter

Zusammenhang des Grammatikerwerbs im Kindergartenalter in Abhängigkeit der familialen Anregungsqualität und des sozialen Hintergrunds

Zusammenhang des Grammatikerwerbs im Kindergartenalter in Abhängigkeit der institutionellen Anregungsbedingungen im Kindergarten

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

**Simultanübersetzungen: Clusterstichprobe bei ausgewählten Veranstaltungen
n = 123 (2005)**

Unterschiedliche Bedeutung von Qualitätskriterien von Simultanübersetzungen
bei Mediziner:innen (46), Techniker:innen (29) und Diplomat:innen (48)

Qualitätskriterien von Simultanübersetzungen:

Inhalt

inhaltliche Konsistenz
logische Geschlossenheit
Vollständigkeit der Information

Korrektheit

korrekte Grammatik
korrekte Terminologie
passender Stil

Ausdruck

flüssige Wiedergabe
angenehme Stimme
native Aussprache

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

Kopfschmerz: Feldexperiment - Zeitlich begrenzte Befragung und Beobachtung
in einer Kopfschmerzambulanz des AKH Wien
Untersuchte Gespräche n = 75, genauere Analyse n = 37 (2011)

Zusammenhang von Schmerzdarstellung und Krankheitserzählungen bei
gedolmetschten und nicht gedolmetschten Gesprächen zu Kopfschmerzen

Bedingungen:

personenbezogen

Erstsprache Deutsch – Türkisch

Erstgespräch – Zweitgespräch

Geschlecht PatientIn

Geschlecht ÄrztIn

gesprächsbezogen

Gesprächsdauer

Dauer bis zur ersten
redestrukturierenden

Äußerung

Externe Unterbrechungen
(Anzahl und Art)

Anzahl der Wörter

dolmetschbezogen

Kein Dolmetsch

Laien-Dolmetsch

Profi-Dolmetsch

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

Beipacktexte: Laborexperiment von Studierenden
Beispielhafte Beipacktexte mit unterschiedlicher Textqualität (2011)

Auswirkungen der Textqualität (Gestaltung, Formulierung)
von Beipacktexten für rezeptfreie Medikamente auf deren Verständlichkeit

- Wissenstest – Wurden die wichtigsten Inhalte vermittelt?
- Beurteilung der Textqualität (verständlich, lesbar, hilfreich....)
- Verbesserungsvorschläge

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

E-Mails versus geschriebene Briefe: Korpuslinguistik (fiktiv)

Kommt es in E-Mails häufiger zu Fehlern als in mit der Hand geschriebenen Briefen? Welche Arten von Fehlern sind für welche Textform charakteristisch?

→ Zufällig ausgewählte E-mails und Briefe einer bestimmten Kategorie
z.B. Beschwerden oder Leser:innenbriefe

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

Dialekt an der Uni?

Einstellungen von Studierenden zum variationspezifischen Gebrauch der deutschen Sprache im institutionellen Umfeld universitärer Lehre
(Dissertation Lisa Krammer 2022)

Wird Dialekt an der Uni gesprochen, wenn ja, von wem, in welchem Kontext und in welchem Ausmaß? Was halten Studierende und Lehrende davon?

<i>quantitativ</i>	<i>qualitativ</i>	<i>Quellenstudium</i>
 Online-Fragebogen (n= 500)	 Leitfaden- interviews	Universitätsgesetz, Statuten, Entwicklungspläne, Leistungsberichte, Studienrecht, ...
Studierende aus 5 Studienrichtungen	Lehrende	 Dokumente

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

Fehlererkennung und Fehlertoleranz:

Befragung von Studierenden aus Frankreich (Paris n= 197)
und Italien (Bologna n= 78) (Habil Holger Wöchele 2019)

Von welchen Merkmalen hängt es ab, dass Fehler als Fehler erkannt werden?
Wie irritierend werden erkannte Fehler erlebt?
Spielt dabei die Erstsprache eine Rolle?
Spielt dabei die Art des Fehlers eine Rolle?

Anhand von französischen/italienischen Beispielsätzen mit eingebauten Fehlern wird untersucht, ob die Fehler erkannt werden, und wenn ja, wie sehr diese als irritierend erlebt werden.

Arten von Fehlern:

Kongruenz, Pronomen, Präpositionen, Artikel, Orthographie, Morphologie....

Sprachwissenschaftliche Fragestellungen zur quantitativen Erforschung

Lernstile:

Befragung von Studierenden aus verschiedenen wirtschaftswissenschaftlichen Studienrichtungen in Wien (n= 198) (WU 2017)

Welche Lernstile haben Studierende?

Sind diese abhängig von der Studienrichtung?

Zeigen sich Unterschiede nach Sprachkenntnissen oder sozialen Merkmalen?

Entspricht die Selbsteinschätzung dem Testergebnis?

Anhand von 80 dichotomen Einstellungsfragen

z.B. „Ich sage klar und ohne Umschweife was ich mir denke.“ ja/nein

wird der Lernstil der Person ermittelt.

4 Lernstile: Aktiv, Reflexiv, Theoretisch, Pragmatisch

Item-Formulierung Antwortverhalten

Zustimmung Ablehnung

Beide, Mann und Frau sollten zum Haushaltseinkommen beitragen.	unbestimmt versus konkret	60,1%	11,2%
Am besten ist es, wenn beide, Mann und Frau, halbtags berufstätig sind und sich gleich um Haushalt und Kinder kümmern.		30,4%	24,2%
Frauen sind von Natur aus besser dazu geeignet, Kinder aufzuziehen.	„Allgemeinplatz“	68,2%	13,8%
Berufstätigkeit ist der beste Weg für eine Frau, um unabhängig zu sein.		70,0%	10,5%
Frauenemanzipation ist eine notwendige und gute Entwicklung.	„Reizwort“ versus „Neutralwort“	55,3%	15,9%
Gleichstellung der Frau und Abbau von Diskriminierung von Frauen sind wichtig.		78,0%	3,0%

Zustimmung Ablehnung

Es gibt einen Gott, der den Lauf der Welt beeinflusst.	62,7%	37,3%
Das Leben wird bestimmt durch die Gesetze der Natur.	78,7%	21,3%

4er-Skala

Anzahl

		f70_16 Es gibt einen Gott, der den Lauf der Welt beeinflusst				Ges
		1 stimme sehr zu	2	3	4 stimme gar nicht zu	
f70_5 Das Leben wird bestimmt durch die Gesetze der Natur	1 stimme sehr zu	125	126	83	121	455
	2	126	210	116	69	521
	3	70	75	46	22	213
	4 stimme gar nicht zu	46	11	8	8	73
Gesamt		367	422	253	220	1262

47%

31%

7%

16%

Zustimmung Ablehnung

Ein Kind wird darunter leiden, wenn die Mutter berufstätig ist.	65,4%	19,0%
Eine berufstätige Frau kann ihrem Kind genauso viel Wärme geben, wie eine Frau, die nicht arbeitet.	46,0%	35,4%

5er-Skala (mit neutraler Kategorie)

23%

Anzahl	f52_8 Ein Kleinkind wird darunter leiden, wenn die Mutter berufstätig ist					Ges	
	1 stimme voll zu	2	3	4	5 lehne voll ab		
f52_4 Eine berufstätige Frau kann ihrem Kind genauso viel Wärme geben wie eine Frau, die nicht arbeitet.	1 stimme voll zu	87	71	62	55	94	369
	2	63	111	51	44	19	288
	3	94	74	79	16	7	270
	4	146	89	27	18	5	285
	5 lehne voll ab	175	28	11	6	14	234
Gesamt		565	373	230	139	139	1446

3%

Likertskala

Ziel: Möglichst breite Streuung über die verwendeten Kategorien erreichen, „Wischi-Waschi-Items“ verhindern, möglichst wenig Anhäufung in einer/wenigen Kategorien

Vorgansweise:

Relevante Items formulieren und summieren. Aus dem Summenindex Quartile bilden – das niedrigste und das höchste Quartil auswählen – Verwendete Items zwischen diesen beiden Gruppen auf signifikante Mittelwertunterschiede testen.

Wie hoch ist der Mittelwertunterschied – und ist er signifikant?

Wenn nicht – Item ausschließen.

Erhebung: fessel Institut 2000, Stichprobe: n = 1500 (Öst. Bev.)

Zustimmung

Es gibt nichts Schlechtes, an dem nicht auch etwas Gutes wäre.	86,0%
Es ist für mich wichtig im Einklang der Natur zu leben.	93,6%
Die Liebe zu einem Menschen ist das Wichtigste im Leben.	94,0%
Meine Familie bedeutet mir alles.	94,2%
Die Arbeit bedeutet mir alles.	47,1%

„Wischi-
Waschi“

Guttman-Skala: „Niederschwellig“ im Übergang zu
„Höberschwellig“ – Theorie überprüfen!
Ausmaß der Zustimmung als **Indikator für „Normalität“**

Ausländer sollten ihren Lebensstil ein bisschen besser an den der Inländer anpassen.	61%
Man sollte Ausländern politische Betätigung im Inland untersagen.	30%
Wenn die Arbeitsplätze knapp werden, sollte man die Ausländer wieder in ihre Heimat zurückschicken.	25%
Ausländer sollten sich ihre Ehepartner unter ihren eigenen Landsleuten suchen.	24%

Guttmankala

Ziel:

Anhand einer Theorie soll „Nieder-“ und „Höher-schwelligkeit“ der Zustimmung zu Items in eine Reihenfolge gebracht werden. Wie gut dies gelingt, gibt der **Reproduktionskoeffizient** an.

Vorgansweise: Beispiel (aus Wikipedia)

- A) Sind Sie bereit, Moslems zu ermöglichen, in Ihrem Land zu leben?
- B) Sind Sie bereit, Moslems zu ermöglichen, an Ihrem Wohnort zu leben?
- C) Sind Sie bereit, Moslems zu ermöglichen, in Ihrem Wohnviertel zu leben?
- D) Sind Sie bereit, Moslems zu ermöglichen, direkt neben Ihnen zu leben?
- E) Würden Sie Ihrem Kind ermöglichen, einen Moslem zu heiraten?

Prinzip: Wer Frage C mit Ja beantwortet, sollte auch Fragen A und B bejahen.

	A Land	B Ort	C Viertel	D Haus	E Bett
1	x	x	x	x	x
2	✓	x	x	x	x
3	✓	✓	x	x	x
4	✓	✓	x	x	x
5	✓	✓	✓	x	x
6	✓	✓	✓	x	x
7	✓	✓	✓	x	x
8	✓	✓	✓	x	x
9	✓	✓	✓	✓	x
10	✓	✓	✓	✓	x
11	✓	✓	✓	✓	✓
12	✓	x	✓	x	x
13	✓	✓	✓	x	x
14	✓	✓	✓	✓	x
15	x	x	x	x	✓

Guttmanuskala

Reproduktionskoeffizient

= Erwartungsgemäße

Antworten / Alle Antworten

(= Items x Fälle)

Hier: Items = 5, Fälle = 15

erwartungsgemäße Antworten = 65

Alle Antworten = 75

$$65 / 75 = 0,87$$

Guttman-Skala:

Wenn zumindest 0,9

Semantisches Differenzial

Ziel: Untersuchen von affektiven Wortbedeutungen (Einstellungsforschung)

Welche Eigenschaften und Vorstellungen verbinden Sie mit dem Glauben an einen Gott bzw. an höhere Wesen und Mächte?

INT.: KARTE VORLEGEN; EINZELNABFRAGEN!

	LINKEEigen- schaft trifft voll u.ganz zu					RECHTEEigen- schaft trifft voll und ganz zu	
	1	2	3	4	5	6	
gütig-vergebend	1	2	3	4	5	6	rächend
zerstörerisch	1	2	3	4	5	6	schöpferisch
väterlich	1	2	3	4	5	6	mütterlich
allmächtig	1	2	3	4	5	6	ohnmächtig
tröstend-nah	1	2	3	4	5	6	unbeteiligt-fern
einengend	1	2	3	4	5	6	befreiend
sexfeindlich	1	2	3	4	5	6	sexfreundlich
ängstigend	1	2	3	4	5	6	beruhigend
für die Armen da	1	2	3	4	5	6	für die Reichen da
streng	1	2	3	4	5	6	liebvoll

Vorgansweise:

Statt Zustimmung zu ausformulierten Items („*Gott lenkt den Lauf der Welt*“)

werden Polaritätsprofile – Gegensatzpaare formuliert

– anschl. dimensionierende Verfahren zeigen Bedeutungsdimensionen

richtig							falsch	
dumm		Freundschaft						klug
oft	0	0	0	0	0	0	selten	
freundlich	0	Wahrheit					0	feindlich
verantwortungsvoll	0					0	verantwortungslos	
geschickt	0	Rücksichtslosigkeit					0	ungeschickt
aktiv	0					0	passiv	
einfach	0	Lüge					0	schwierig
egoistisch	0					0	selbstlos	
anziehend	0					0	abstoßend	
männlich	0	0	0	0	0	0	weiblich	
notwendig	0	Irrtum					0	nebensächlich
destruktiv	0					0	konstruktiv	
absichtlich	0	Aufrichtigkeit					0	unabsichtlich
moralisch	0					0	unmoralisch	
schmutzig	0					0	sauber	
mutig	0	0	0	0	0	0	feige	
hart	0	0	0	0	0	0	weich	
stark	0	0	0	0	0	0	schwach	
akzeptabel	0	0	0	0	0	0	inakzeptabel	
warm	0	0	0	0	0	0	kalt	

<p><i>„Das Wichtigste, was Kinder lernen müssen, ist Gehorsam.“</i></p>	Gesamt 41%	
	Männer 42%	Frauen 41%
	unter 30 J. 29%	über 60 J. 65%

<p><i>„Das Wichtigste, was Kinder lernen müssen, ist Teilen.“</i></p>	Gesamt 81%	
	Männer 76%	Frauen 84%
	unter 30 J. 74%	über 60 J. 90%

<p><i>„Ich bin der Meinung, Frauen sind ohnehin gleichgestellt.“</i></p>	<p>Gesamt 35%</p>	
	<p>Männer 43%</p>	<p>Frauen 29%</p>
	<p>über 60 J. 42%</p>	<p>unter 30 J. 28%</p>
	<p>Männer über 60 J. 50%</p>	<p>Frauen unter 30 J. 20%</p>

<p><i>„Get rid of parliament and elections, have a strong leader to decide.“</i></p> <p><i>„Von Zeit zu Zeit würde ich mir eine Diktatur wünschen, dann gäbe es nicht so viele Missstände.“</i></p>	Gesamt 32%	
	Männer 29%	Frauen 34%
	über 40 J. 26%	unter 40 J. 36%
	„reich“ 12%	„arm“ 43%

	Man kann den meisten Menschen vertrauen.	Man sollte lieber vorsichtig sein.
dichotom	18%	82%
	Vertrauen in die meisten Menschen	Kein Vertrauen in die meisten Menschen
Skala 1 großes Vertrauen 7 gar kein Vertrauen	37% 1 - 2 - 3	25% (4)
		38% 5 - 6 - 7

Neben der Formulierung bestimmt die **Skala** das Antwortverhalten

Rating-Skala: Passendes Spektrum der Skala

Wie lange schauen Sie an einem normalen Wochentag fern?

Version A:

- Bis ½ Std 7,4%
- ½ bis 1 Std 17,7%
- 1 bis 1 ½ Std 26,5%
- 1 ½ bis 2 Std 14,7%
- 2 bis 2 ½ Std 17,7%
- Mehr als 2 ½ Std 16,2%

37% der Befragten schauen mehr als 2 ½ Stunden fern

Version B:

- Bis 2 ½ Std 62,5%
- 2 ½ bis 3 Std 23,4%
- 3 bis 3 ½ Std 7,8%
- 3 ½ bis 4 Std 4,7%
- 4 bis 4 ½ Std 1,6%
- Mehr als 4 Std 0,0%

16 % der Befragten schauen mehr als 2 ½ Stunden fern

Wenn die Befragten darauf angewiesen sind, etwas zu schätzen, sind die vorgegebenen Antwortskalen entscheidend.

Gütekriterien einer Messdimension, eines Items

Item = eine Einzelfrage als Bestandteil einer Messdimension

Eine Frage ist **reliabel** (zuverlässig) und **valide** (gültig) wenn:

- sie verständlich, eindeutig, neutral formuliert ist
- konstante Befragungssituation gegeben ist (Interviewerverhalten...)
- Anforderungen an die **Antwortmöglichkeiten** bzw. **Antwortskala**:
 - symmetrisch** (bei ordinal und metrisch)
 - eindimensional** (kein „und“ und gleiche Pole)
 - disjunkt** (Kategorien überschneiden sich nicht)
 - präzise**
 - erschöpfend** (keine Kategorie fehlt – ev. „Sonstiges“)

Die Frage als Messinstrument - Antwortformate

- Offenes Antwortformat: Anmerkung, Begründung, Sonstiges, u.ä.
- Dichotom: nur zwei Alternativen, z.B. trifft zu / trifft nicht zu
- Set von dichotomen Items = Mehrfachantwort
- Rating: Abgestufte Antwortskala (sehr zufrieden bis sehr unzufrieden)
- Ranking: Einzelitems in eine Reihenfolge bringen (Wichtigstes – Unwichtigstes)

Nominal-Skala

Bsp.: Frage an Frauen nach ATA (außergerichtlichem Tatausgleich) bei häuslicher Gewalt: „Bitte beschreiben Sie Ihre Gewaltgeschichte – was trifft auf Sie zu?“

Präzise? Erschöpfend? Disjunkt? Eindimensional?

a9 Gewaltgeschichte

		Häufigkeit	Prozent
Gültig	1 lange Phase von Übergriffen	47	29,0
	2 kurze Phase von Übergriffen	5	3,1
	3 lange Phase von Übergriffen, habe mich auch gew ehrt	49	30,2
	4 erster Übergriff	56	34,6
	Gesamt	157	96,9
Fehlend	99 keine Angabe	5	3,1
Gesamt		162	100,0

Nominal-Skala

Bsp.: Frage nach der aktuellen Partnersituation an Frauen nach Tatausgleich bei häuslicher Gewalt: „Was trifft auf Ihre aktuelle Partnersituation zu?“

Präzise? Erschöpfend? Disjunkt?

e2 Partnersituation aktuell

		Häufigkeit	Prozent
Gültig	1 gegenseitiger Respekt, weitgehend konfliktfrei	36	22,2
	2 Konflikte, die gut ausgetragen werden können	29	17,9
	3 gelegentlicher Gewaltausbruch	17	10,5
	4 häufiger Gewaltausbruch	3	1,9
	5 leben nebeneinander her, gehen Konflikten aus dem Weg	9	5,6
	6 habe keine Partnerbeziehung	57	35,2
	Gesamt	151	93,2
Fehlend	99 keine Angabe	11	6,8
Gesamt		162	100,0

Antwortskala als „Rating“: mehr als zwei abgestufte Antwortmöglichkeiten

Verbale Bezeichnungen der Antwortmöglichkeiten:

- **Unspezifische verbale Bezeichnung aller Antwortmöglichkeiten: ordinal**
z.B. *sehr / eher / eher nicht / gar nicht*
z.B. *immer / häufig / manchmal / selten / gar nicht*
- **Verbale Bezeichnung der Endpunkte der Skala: meist metrisch**
z.B. Arbeitsbelastung von 0 (= *gar nicht*) bis 10 (= *sehr stark*)

Art der Polarität:

- **Unipolar:** von der Abwesenheit bis zum höchstem Ausmaß des Merkmals
z.B. *Angst: gar keine - sehr große*
- **Bipolar:** von negativer - über neutrale - zu positiver Bewertung
z.B. Zustimmung zu einer Aussage:
stimme sehr zu / weder noch / stimme gar nicht zu

Unterschiedliche Differenzierung:

- **gering:** z.B. Arbeitsbelastung: *stark / gering / gar nicht*
- **hoch:** z.B. Arbeitsbelastung von 0 (= *gar nicht*) bis 10 (= *sehr stark*)

Zur Verwendung von einer neutralen Mittelkategorie

Ungerade Antwortskala: Vorhandensein einer **Mittelkategorie**

Vorteil: Weniger Antwortverweigerung aufgrund Ausweichmöglichkeit

Nachteil: Vorhandensein neutraler Antworten (keine (pos./neg.) Bewertung)

Gerade Antwortskala: keine **neutrale Mittelkategorie** („forced choice“)

Vorteil: bei jeder gültigen Antwort ist eine pos./neg. Bewertung vorhanden

Nachteil: Mehr fehlende Antworten von jenen, die sich nicht entscheiden wollen

Verwendung je nach Fragestellung

z.B. Eignungstests: meist keine neutrale Mittelkategorie

Wenn möglich, **Antwortkategorie „weiß nicht“** vermeiden;

- dient ebenfalls zur Antwortverweigerung;

„Alles in allem, wie zufrieden sind Sie mit Ihrem Leben insgesamt?“

Skala 1	%	%	Skala 2
sehr zufrieden	17	17	sehr zufrieden
eher zufrieden	62	52	eher zufrieden
	-	18	weder noch
eher unzufrieden	13	9	eher unzufrieden
sehr unzufrieden	3	3	sehr unzufrieden
weiß nicht	5	1	weiß nicht
	100%	100%	

„Alles in allem, wie zufrieden sind Sie mit Ihrem Leben insgesamt?“

Frage 3	%	%	Frage 4
„zufrieden“ 1	22	18	0 „zufrieden“
2	25	17	1
3	18	22	2
4	9	13	3
5	8	7	4
	-	7	5
6	4	4	6
7	6	4	7
8	4	4	8
9	2	1	9
„unzufrieden“ 10	2	3	10 „unzufrieden“
	100%	100%	

Gegenüberstellung in Kreuztabelle – 10er versus 11er-Skala

„zufrieden“

	1	2	3	4	5	6	7	8	9	10	Gesamt
0	92	7	2	1							102
1	19	59	4						2		84
2	1	47	40	1	1	1			1		92
3	1	9	30	14	2		1	1		1	59
4	1	1	3	22	5	4					36
5		1	1	1	21	8	4		1	1	38
6					5	8	4	1			18
7		1			2	2	10	5	1	1	22
8					3	2	1	8	4		18
9	1	1						4	4	3	13
10					1			1		13	15
Gesamt	115	126	80	39	40	25	20	20	13	19	497

Gegenüberstellung in Kreuztabelle – nach einem „Richtungswechsel“

„unzufrieden“

„zufrieden“

	1	2	3	4	5	6	7	8	9	10	Gesamt
1	5	4	2		2	2		2	2	10	29
2	7	17	2	1	3			2	3	2	37
3	2	6	17	3	3	2	6	3	2	1	45
4	1	3	4	13	3	2	7	3			36
5	4	3	4	4	10	4	2	1		1	33
6	1	1		2	8	8	2	2	1		25
7		3	7	11	4	5	3	2		1	36
8	9	22	33	3	4	2		4	3		80
9	17	54	10	2	1			1	1		86
10	69	13	1		2				1	4	90
Gesamt	115	126	80	39	40	25	20	20	13	19	497

Fragebogenkonstruktion



Sozialwissenschaftliche Befragung

Verbale Stimuli (Fragen) → verbale Reaktionen (Antworten)

Es werden individuelle Daten gesammelt und bei der Auswertung zu Gruppen oder Typen zusammengefasst. Zweck ist die Verallgemeinerung von Gesetzmäßigkeiten und Gruppenunterschieden.

Es wird nicht soziales Verhalten erfasst, sondern lediglich die Reaktion auf verbale Stimuli (= verbales Verhalten) - also die Reaktion auf Frageformulierungen und Antwortskalen.

Kommunikationsart: face-to-face, schriftlich, telefonisch

Häufigkeit: einmalig - Querschnitt oder mehrmals - Längsschnitt

- Trendanalyse: gleicher Fragebogen – anderer Zeitpunkt
- Panelanalyse: gleiche Gruppe mehrmals befragt (aufwändig)

Standardisierte Befragung: Befragung mittels Fragebogen

Soziolinguistik: Thema Sprachwandeltendenzen: untersucht wird der Sprachgebrauch:

- apparent time Design: zwei Altersgruppen werden zur gleichen Zeit befragt
- real time Design: die gleiche Altersgruppe wird zu zwei Zeitpunkten befragt

Erhebungsmodi für Befragungen

	Vorteil	Nachteil
Face-to-face-Befragung	InterviewerInnen befragen die Personen vor Ort; Umfangreiche Fragebögen möglich Sorgfältige Erhebung und daher sehr gute Datenqualität	teuer! Gefahr des Interviewer-Effekt bei nicht ausreichender Interviewer-Schulung (neutrales Verhalten!)
Schriftliche Befragung	Relativ kostengünstig können große Zielgruppen befragt werden, ist besonders gut geeignet, wenn die Adressen der Zielgruppe vorliegen.	Geringe Rücklaufquote bei mangelnder Motivation zur Teilnahme
Telefonische Befragung	Sehr kostengünstig und rasche Durchführung Befragte müssen vorher schriftlich informiert werden	Nur für kurze Fragebögen geeignet, weil Gefahr, dass das Gespräch abgebrochen wird
Online-Befragung	Kostenlos, rasch Liste mit Mailadressen und geeignetes Online-Befragungstool muss vorhanden sein (SurveyMonkey, Soci-survey, Limesurvey...)	Nur bei bestimmten Zielgruppen geeignet, beispielsweise RepräsentantInnen von sozialen Einrichtungen, sozialen Unternehmen u.ä. Ansonsten: Problem der mangelnden Repräsentativität!

Grundsätze zur Fragebogenkonstruktion

Voraussetzungen:

- Präzise Formulierung der Fragestellung
- sorgfältige Operationalisierung mithilfe von Theorie und eigener Erfahrung

Passende Fragen wählen:

Offene Fragen: explorativ. Es empfiehlt sich, am Ende eines FBs eine offene Frage zu stellen – Hinweise, was bei der Operationalisierung übersehen wurde

Geschlossene Fragen:

vorgegebene verbale Antwortkategorien oder Boxen zum ankreuzen

Kriterien der Frageformulierung müssen erfüllt sein:

einfache, allgemeinverständliche Wortwahl

keine Fachausdrücke, keine Reizwörter, keine Abkürzungen

keine Suggestion (z.B. Argumente voranstellen bewirkt Antwortverzerrungen)

keine hypothetischen Fragen

Verneinung vermeiden = wenn möglich positive Formulierungen

Fragebogendramaturgie

Einleitungstext:

- Anrede auf die befragte Gruppe abstimmen
- Kurze Vorstellung von Fragestellung, Zweck, Auftraggeber und Durchführende/r
- Kurze Anleitung zum Ausfüllen mit Zeitangabe
- Bitte um Aufrichtigkeit, Vollständigkeit und Zusicherung der Anonymität
- Danksagung für die Mitarbeit

Fragebogen:

- zu Beginn Interesse für das Thema wecken, möglichst keine langwierigen Fragebatterien am Anfang, leichten **Einstieg** schaffen
- FB übersichtlich in **Abschnitte** gliedern, mit Überschriften
- chronologisch passende **Abfolge** (keine zu weiten Themensprünge)
- **Filterfragen** verwenden: z.B. „Sind Sie berufstätig“ → falls nein, weiter bei Frage 4
- sensible und schwierige Fragen eher ans Ende stellen
- demografische Fragen ans Ende stellen
- Tricks: z.B. Einkommen nicht als offene Frage sondern Kategorien vorgeben
z.B. Nicht direkt nach Parteizugehörigkeit fragen,
sondern nach dem Grad der Zustimmung zum Parteiprogramm

Die elf Gebote bei der Fragebogenkonstruktion

1. Allgemeinverständlich formulieren!
2. Keine zu komplizierten Fragetexte!
3. Immer nur ein Sachverhalt pro Frage! (Eindimensionalität)
4. Keine hypothetischen Fragen!
5. Möglichst keine Verneinungen!
6. Keine Suggestionen!
7. Konkrete zeitliche Bezüge angeben!
8. Antworten müssen zur Frage passen!
9. Keine Antwortkategorie darf fehlen (erschöpfend) und:
die Kategorien dürfen sich nicht überschneiden! (disjunkt)
10. Achte auf die Platzierung der Frage im Instrument!
11. Mach einen Pretest!

Fehlerquellen bei Befragungen

Die Befragten:

- sozial erwünschtes Antwortverhalten
- Meinungslosigkeit („*weiss nicht*“ oder Ja-sage-Tendenz)
- Unaufmerksamkeit (zu langer Fragebogen, mangelnde Motivation...)
- Antwortverweigerung (abhängig von Thema, Formulierung, verbalen Fähigkeiten der Befragten...)

Die Fragen:

- Frageformulierung bzw. Position der Frage im Fragebogen
- Beeinflussung durch Antwortskala bzw. Antwortkategorien

Die Interviewsituation:

- Spezifische Reaktion auf InterviewerInnen: Alter, Geschlecht, Kleidung, Verhalten (Mimik, Gestik), Sprache..... Auswirkungen hängen von der Frage ab
- Anwesenheit dritter Personen, Gruppeneffekte (z.B. Befragung in Schulklassen)

**Vergleichbarkeit von erhobenen Daten ist nur dann gegeben wenn:
Frageformulierung, Antwortmöglichkeiten, Reihenfolge der Fragen,
Interviewsituation in gleicher Weise (standardisiert) erfolgen.**

Beispiel Fehlerquelle „sozial erwünschtes Antwortverhalten“

Eine britische Untersuchung zum Sexualverhalten in den 90iger Jahren fragte nach der Anzahl der Sexualpartner:innen.

Frauen hatten durchschnittlich 2,9 Sexualpartner.

Männer hatten durchschnittlich 11 Sexualpartnerinnen.

„It takes two to tango.....“



Besonders bei unangenehmen/intimen Fragen spielt soziale Erwünschtheit eine große Rolle.

Beispiel Fehlerquelle „sozial erwünschtes Antwortverhalten“

Eine deutsche Studie fragte nach dem Bekanntheitsgrad des Staatssekretärs Dieter Köstritz.

- 8% aller Befragten kannten ihn.
- 18% aller Befragten mit höherem Bildungsabschluss kannten ihn.

Dieter Köstritz existiert nicht.



Nichtwissen wird meist sehr ungern eingestanden, auch wenn es die Kategorie „weiß nicht“ gibt.

Beispiel Fehlerquelle Frageformulierung: Umfrage zur Samstagsarbeit

Umfrage A
Auftraggeber:
Gewerkschaft

→ 95% sind gegen
die Samstagsarbeit.

Die Arbeitgeber und manche Politiker wollen vor allem den Samstag wieder zum normalen Arbeitstag machen. Wie wäre das, wenn Sie regelmäßig am Samstag arbeiten müssten?

Würde mir nichts ausmachen

Wäre ein Verlust meiner Lebensqualität

Umfrage B
Auftraggeber:
Industriellenvereinigung

→ 72% würden auch
am Samstag arbeiten.

Inwieweit wären Sie bereit, samstags zu arbeiten, wenn es für die wirtschaftliche Situation Ihres Unternehmens gut wäre?

Gelegentlich, wenn dafür an einem anderen Tag arbeitsfrei ist

Häufiger, an mehreren Samstagen, wenn dafür ein Zusatzurlaub von mehreren Tagen herauskommt

Abwechseln, einmal 6 Tage und in der nächsten Woche 4 Tage arbeiten.

Nein, nicht bereit

Beispiel Fehlerquelle Antwortskala

Wie lange schauen Sie an einem normalen Wochentag fern?

Version A:

- Bis ½ Std 7,4%
- ½ bis 1 Std 17,7%
- 1 bis 1 ½ Std 26,5%
- 1 ½ bis 2 Std 14,7%
- 2 bis 2 ½ Std 17,7%
- Mehr als 2 ½ Std 16,2%

37% der Befragten schauen mehr als 2 ½ Stunden fern

Version B:

- Bis 2 ½ Std 62,5%
- 2 ½ bis 3 Std 23,4%
- 3 bis 3 ½ Std 7,8%
- 3 ½ bis 4 Std 4,7%
- 4 bis 4 ½ Std 1,6%
- Mehr als 4 Std 0,0%

16 % der Befragten schauen mehr als 2 ½ Stunden fern

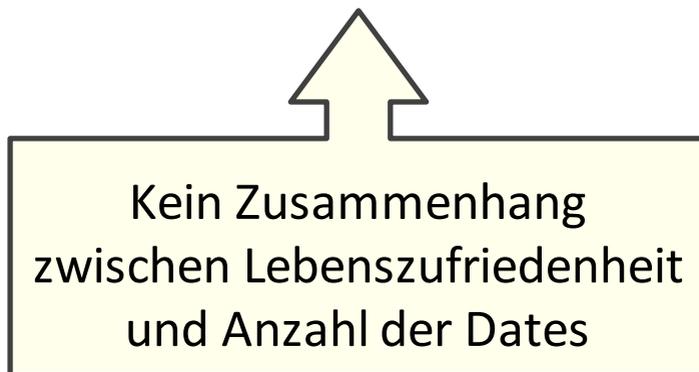
Wenn die Befragten darauf angewiesen sind, etwas zu schätzen, sind die vorgegebenen Antwortskalen entscheidend.

Beispiel Fehlerquelle Fragepositionierung: Lebenszufriedenheit in Abhängigkeit von sexuellem Verhalten

Umfrage unter Studierenden (USA)

Version 1:

- 1.) Alles in allem, wie zufrieden sind Sie mit Ihrem Leben?
- 2.) Wie viele Dates hatten Sie im letzten halben Jahr?



Version 2:

- 1.) Wie viele Dates hatten Sie im letzten halben Jahr?
- 2.) Alles in allem, wie zufrieden sind Sie mit Ihrem Leben?



Beispiel Fehlerquelle Fragepositionierung

Retrospektive Befragung von Patient:innen
nach Caring-Bedürfnissen im Krankenhaus: 29 Caring-Befürfnisse

→ Frageblock
*„Wie wichtig ist es Ihnen, dass die
folgenden Dinge erfüllt sind?“*

→ Frageblock
*„Wie sehr wurden die folgenden Dinge
erfüllt, als Sie das letzte Mal im
Krankenhaus waren?“*

Welcher Frageblock sollte zuerst gestellt werden?

Beispiel Fehlerquelle Situation und Gruppeneffekte

Befragung in 8 verschiedenen Settings

Zu wie viel Prozent beteiligen Sie sich an der Hausarbeit in Ihrem Haushalt?

	PartnerIn ist anwesend				PartnerIn ist nicht anwesend			
Befragte/r ist	M	M	F	F	M	M	F	F
Interviewer ist	M	F	M	F	M	F	M	F
	1	2	3	4	5	6	7	8
Antworten in %								
0-10%	35,5	45,3	0,7	1,6	24,6	37,3	0,5	1,5
11-25%	24,6	26,6	2,9	3,3	26,9	23,9	1,0	0,8
26-50%	37,0	25,0	17,3	21,3	42,3	34,3	14,1	8,3
51-100%	2,9	3,1	79,1	73,8	6,3	4,5	84,5	89,4
Fallzahl	169	80	143	61	223	86	210	132

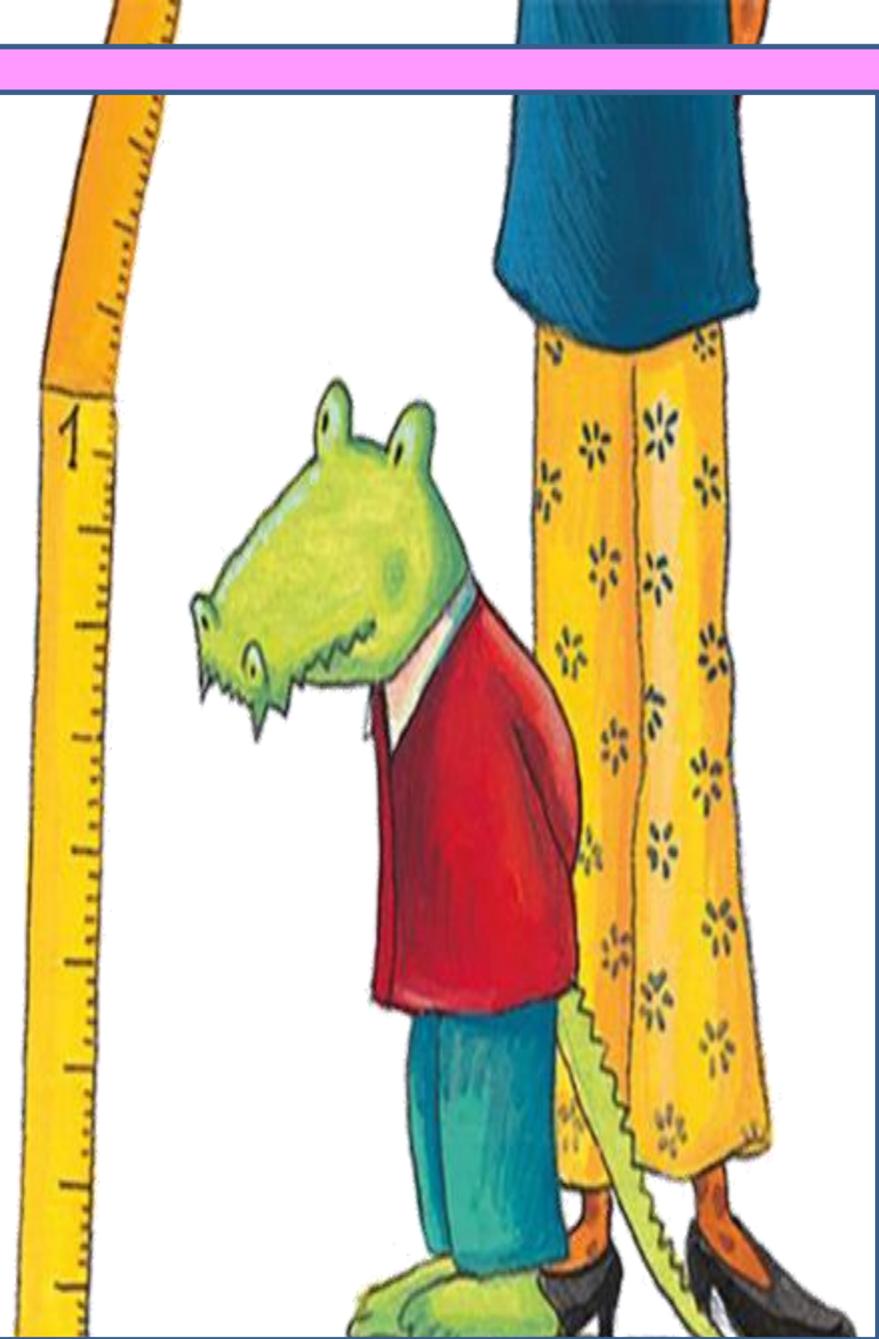
Zu wie viel Prozent beteiligen Sie sich an der Hausarbeit in Ihrem Haushalt?

	Partner:in ist anwesend				Partner:in ist nicht anwesend			
Befragte/r ist	M	M	F	F	M	M	F	F
Interviewer ist	M	F	M	F	M	F	M	F
	1	2	3	4	5	6	7	8
Antworten in %								
0-10%	35,5	45,3	0,7	1,6	24,6	37,3	0,5	1,5
11-25%	24,6	26,6	2,9	3,3	26,9	23,9	1,0	0,8
26-50%	37,0	25,0	17,3	21,3	42,3	34,3	14,1	8,3
51-100%	2,9	3,1	79,1	73,8	6,3	4,5	84,5	89,4
Fallzahl	169	80	143	61	223	86	210	132

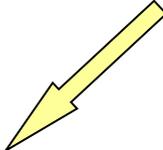
- Männer tendieren dazu, weniger Beteiligung an der Hausarbeit zuzugeben, wenn.....
 - ihre Partnerin anwesend ist
 - die Interviewerin weiblich ist
 - und am meisten, wenn die Partnerin anwesend ist und die Interviewerin weiblich ist
(alle anwesenden Personen sind weiblich)

- Frauen tendieren dazu, mehr Beteiligung an der Hausarbeit zuzugeben, wenn....
 - ihr Partner nicht anwesend ist
 - der Partner nicht anwesend ist und die Interviewerin weiblich ist
(alle anwesenden Personen sind weiblich)

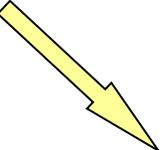
Datenniveaus
= Informations-
gehalt
in den Daten



Skalenniveaus = Informationsgehalt von Daten



Kategoriales Datenniveau



Metrisches Datenniveau

Nominalskala

Kategorien von Eigenschaften, verbal bezeichnet und ohne Rangordnung
(z.B. *Familienstand, Krankheit/Diagnose*)

Intervallskala

Die einzelnen Ausprägungen sind nicht verbal bezeichnet und haben gleiche Abstände
(Punkteskala; z.B. *IQ*)

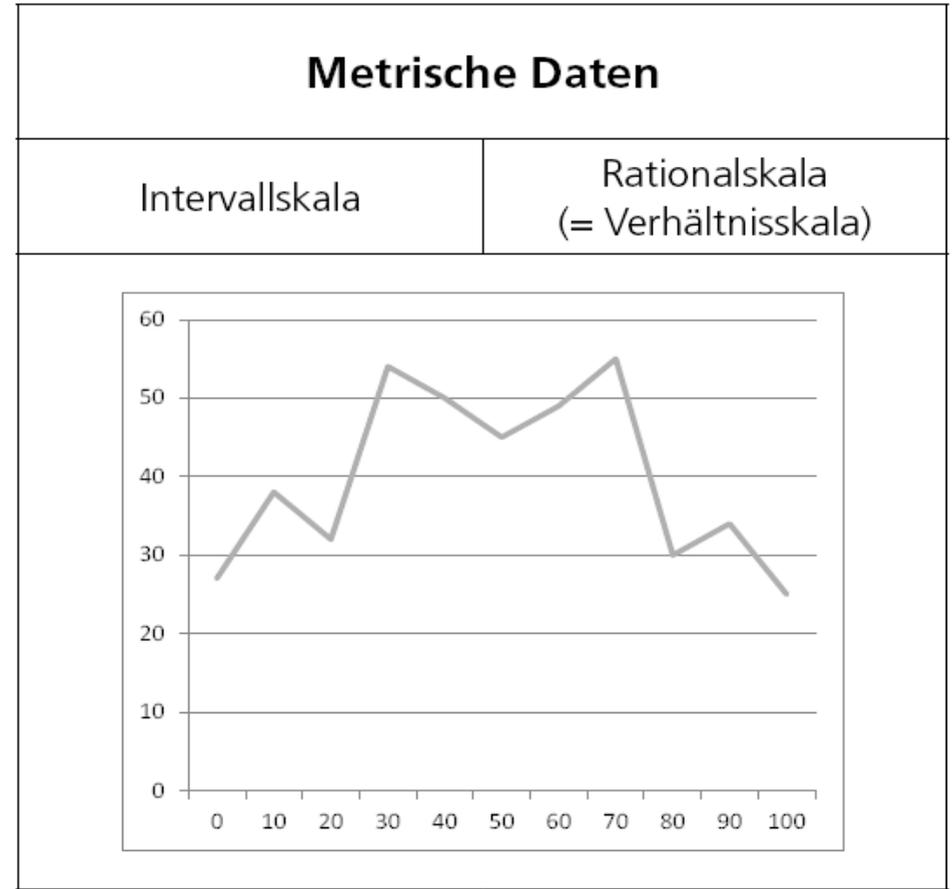
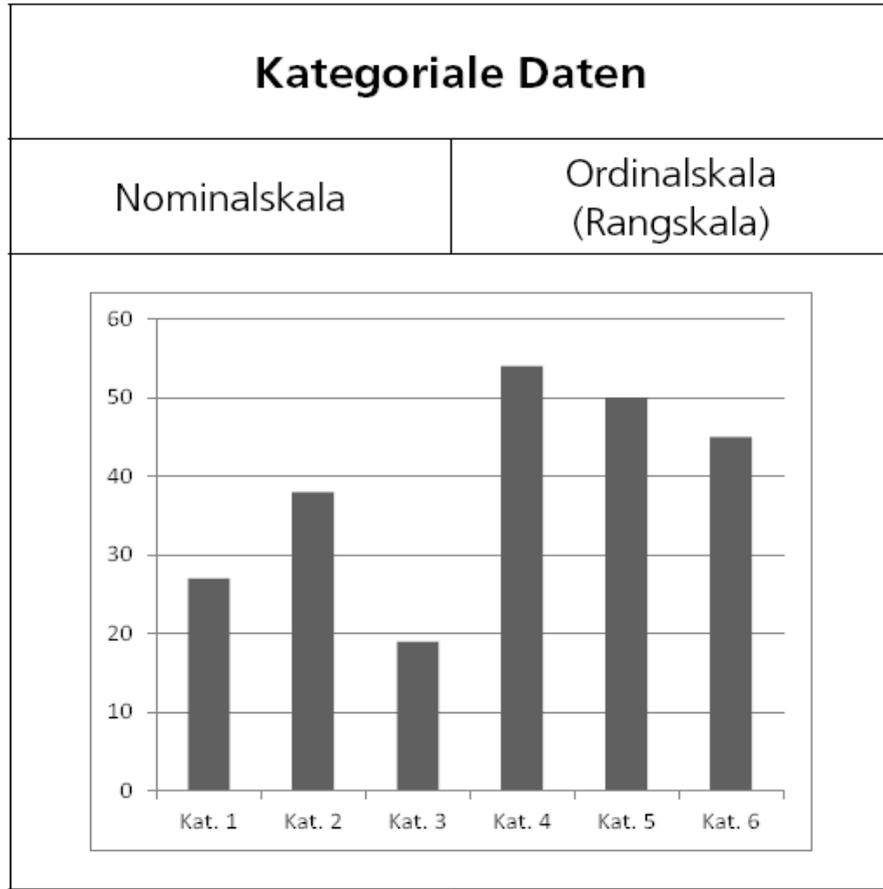
Ordinalskala

Kategorien sind verbal bezeichnet und können in eine Rangordnung gebracht werden
(z.B. *Bildung, Lebenszufriedenheit, Wohlbefinden, Gesundheitszustand in 4 Kategorien*)

Rationalskala

Die Ausprägungen unterliegen einer Messeinheit, haben gleiche Abstände und einen „natürlichen“ Nullpunkt
(z.B. *Alter in Jahren, Körpergröße, Körpergewicht*)

Skalenniveaus Übersicht





Merkmalsausprägungen:

Zuordnung der Werte/Zahlen:

Mathematische Operationen:

Beispiel:

Lebenssituation

- | | | |
|--------------------------|--|---|
| <input type="checkbox"/> | ledig / keine Partnerschaft | 1 |
| <input type="checkbox"/> | in Partnerschaft ohne gemeinsamen Haushalt | 2 |
| <input type="checkbox"/> | in Partnerschaft im gemeinsamen Haushalt | 3 |
| <input type="checkbox"/> | andere Lebensform | 4 |

Nominalskala

Unterscheidung auf Gleichheit =, \neq

Benennung

Häufigkeitsvergleiche

Ordinalskala

Merkmalsausprägungen:

Unterscheidung auf Gleichheit =, ≠
Rangordnung >, <

Zuordnung der Werte/Zahlen:

Rangordnung

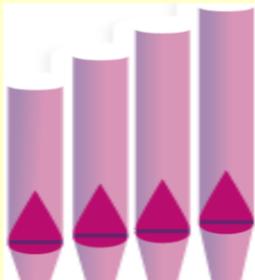
Mathematische Operationen:

Häufigkeitsvergleiche
Rangordnungsrelationen

Beispiel:

Lebenszufriedenheit:

Wenn Sie auf Ihr bisheriges Leben zurückblicken, sind Sie damit...



- sehr zufrieden 1
- eher zufrieden 2
- weder / noch 3
- eher unzufrieden 4
- sehr unzufrieden 5

Intervallskala

Merkmalsausprägungen:

Unterscheidung auf Gleichheit =, ≠

Rangordnung >, <

Intervalle sind gleich groß

Zuordnung der Werte/Zahlen:

Zahlen mit gleicher Differenz

ohne natürlichem Nullpunkt

Mathematische Operationen:

Häufigkeitsvergleiche

Rangordnungsrelationen

Intervallrelationen

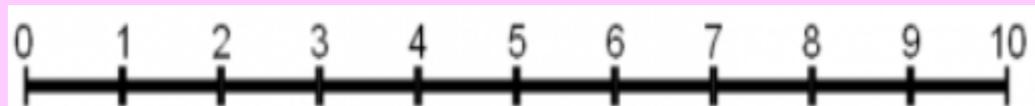
(Addition, Subtraktion)

Beispiel:

standardisierte IQ-Punkte

Schmerzskala 0-10

diverse Indizes





Merkmalsausprägungen:

Zuordnung der Werte/Zahlen:

Mathematische Operationen:

Beispiel:

Alter: Geben Sie bitte Ihre Alter in Jahren an: _____

Rationalskala oder Verhältnisskala

Unterscheidung auf Gleichheit =, ≠

Rangordnung >, <

Intervalle sind gleich groß

natürlicher (absoluter) Nullpunkt

**Messwert „Null“ entspricht Abwesenheit
des Merkmals**

Häufigkeitsvergleiche

Rangordnungsrelationen

Intervallrelationen

Verhältnisrelationen

(Multiplikation, Division)

Je höher das Niveau, desto mehr Informationsgehalt, desto differenziertere Verfahren sind möglich

Messeinheit in gleichen Abständen

Punkteskala

*Nullpunkt
kg, Jahre, ...*

Gruppen
*(ledig - verheiratet -
gesch. - verwitwet)*

$= \neq$

nominal

Reihenfolge
*(sehr - eher -
wenig - gar nicht)*

$\geq \leq$

$= \neq$

ordinal

$- +$

$\geq \leq$

$= \neq$

intervall

$\cdot \div$

$- +$

$\geq \leq$

$= \neq$

rational

metrisch

Informationsgehalt / Analysemöglichkeiten



Fragen in verschiedenen Datenniveaus

Alter der Befragten

NOMINAL

Sind Sie Jugendliche/r
oder Erwachsene/r?

- Jugendliche /r
(bis 19 Jahre)
- Erwachsene /r
(19 Jahre und älter)

ORDINAL

Wie alt sind Sie?

- bis 30 Jahre
- 31-50 Jahre
- 51-70 Jahre
- über 70 Jahre

METRISCH

Wie alt sind Sie?

_____ Jahre

Fragen in verschiedenen Datenniveaus

Musikkonsumgewohnheiten der Befragten

NOMINAL

Wann hören Sie meistens aufmerksam Musik?

- daheim, allein
- daheim, mit anderen
- bei Konzerten
- anderes, und zwar?

ORDINAL

Wie oft hören Sie aufmerksam Musik?

- täglich
- mehrmals pro Wo
- mehrmals pro Mo
- mehrmals pro Jahr
- seltener
- nie

METRISCH

Wie viele Stunden in der Woche hören Sie aufmerksam Musik?

_____ Stunden

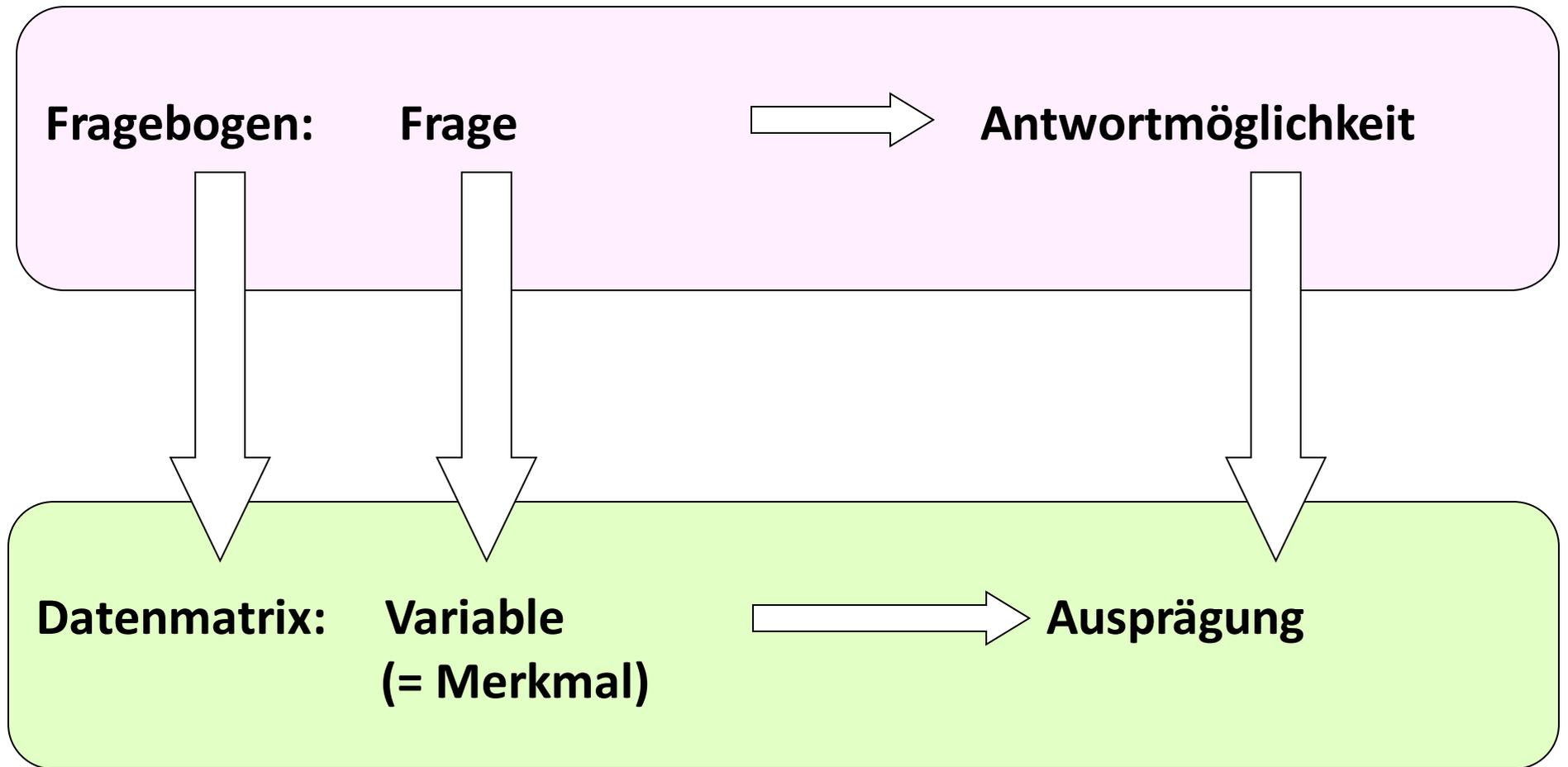
Skalenniveaus Übersicht

Mess-niveau	Eigenschaften	Beschreibung	Beispiele
Nominal	gleich/ungleich	Messwerte gleich oder ungleich	Familienstand, Geschlecht, ...
Ordinal	größer/kleiner	Messwerte lassen sich der Größe nach ordnen; Abstände zw. den Messwerten sind NICHT messbar bzw. nicht gleich	Bildung, Schicht, ...
Intervall	gleiche Abstände (+/-)	gleiche Abstände zw. den Messwerten; Abstände sind messbar	IQ, versch. Punkteskalen, ...
Rational	gleiche Abstände, Nullpunkt und Messeinheit (*:/)	gleiche Abstände zw. den Messwerten; Abstände sind messbar, Messeinheit, Messwertverhältnisse sind berechenbar	Alter, Gewicht, ...

	kategorial		metrisch	
	nominal	ordinal	intervall	rational
Zuordnung	Kategorien sind verbal benannt	Kategorien sind verbal benannt und nach Größe/Ausmaß gereiht; Abstände zwischen den Kategorien sind NICHT MESSBAR	breite Punkteskala, nur die Endpunkte bezeichnet; Oder: Standardisierte Skala aus mehreren Items zusammengesetzt Abstände zwischen den Messwerten sind GLEICH	Messeinheit vorgegeben (Jahre, Stunden, Kilo...) Absolut: Anzahl (Stück)
Messwerte	gleich – ungleich	gleich – ungleich mehr – weniger	gleich – ungleich mehr – weniger gleiche Abstände	gleich – ungleich mehr – weniger gleiche Abstände „natürlicher“ Nullpunkt
Operationen	Häufigkeitsvergl.	Häufigkeitsvergleich Rangordnungsrelation	Häufigkeitsvergleich Rangordnungsrelation Intervallrelation	Häufigkeitsvergleich Rangordnungsrelation Intervallrelation Verhältnisrelation
Transformation	gleich/ungleich	gleich/ungleich größer/kleiner als	gleich/ungleich größer/kleiner als addieren, subtrahieren	gleich/ungleich größer/kleiner als addieren, subtrahieren multiplizieren/dividieren
Beispiel	Familienstand Lebenssituation Geschlecht Herkunftsland	Zufriedenheit 1-5 Schulnoten, Bildung Alkoholkonsum (unspezifisch)	Zufriedenheit 0-10 IQ-Skala Berufsprestigescore	Alter, Einkommen, Reaktionszeit, Anzahl Stunden pro Tag im Internet
erlaubte Aussage	<i>Martin ist ledig, Josefine ist geschieden. .</i>	<i>Martin trinkt mehrmals pro Woche Alkohol, Josefine seltener.</i>	<i>Martins IQ-Wert beträgt 120 Punkte, Josefine hat um 20 Punkte weniger, nämlich 100 Punkte.</i>	<i>Martin ist halb so alt wie Josefine. dafür verdient er viermal soviel wie sie....</i>

**Vom Fragebogen zur Datenmatrix
Datenaufbereitung
Codierung**

Vom Fragebogen zur Datenmatrix



Vom Fragebogen zur Datenmatrix

Wording

„Frage“ (Fragebogen) = „Variable“ (Datenmatrix)

„Antwortmöglichkeit“ (Fragebogen)
= „Ausprägung“ (Datenmatrix)

A Zusammenfassung zur Person und zur Person Ihrer Person Ihres Angehörigen									
1	Sie sind ...	① männlich	② weiblich						
2	Bitte geben Sie Ihr Alter an:	Jahre						
3	Wie lange lebte Ihr/e Angehörige/r in diesem Heim?	Jahre und	Monate					
4	Auf welche Pflegestufe war sie/er eingestuft?	①	②	③	④	⑤	⑥	⑦	⑧ keine
B Ich hatte ausreichend Kontakt ...									
			stimme voll zu		stimme nicht zu				
1	... zum Pflegepersonal	①	②	③	④	⑤	⑥		
2	... zu den Ärztinnen und Ärzten	①	②	③	④	⑤	⑥		
3	... zum Personal betreffend Therapie und Seniorenbetreuung	①	②	③	④	⑤	⑥		

Vom Fragebogen zur Datenmatrix - Layout

Variablen-
name

Code/Wert/
Ausprägung

Wertelabel

Variablen-
label

A Zu Ihrer Person und zur Person Ihrer/Ihres Angehörigen

a1

1 Sie sind ... ① männlich ② weiblich

Geschlecht

a2

2 Bitte geben Sie Ihr Alter an: Jahre

Alter

a3

3 Wie lange lebte Ihr/e Angehörige/r in diesem Heim? Jahre und Monat

Aufenthaltsdauer

a4

4 Auf welche **Pflegestufe** war sie/er eingestuft? ① ② ③ ④ ⑤ ⑥ ⑦ ① keine

Pflegestufe

B Ich hatte ausreichend Kontakt ...

stimme voll zu

stimme nicht zu

b1

1 ... zum Pflegepersonal ① ② ③ ④

Kontakt zu Pflegepersonal

b2

2 ... zu den Ärztinnen und Ärzten ① ② ③ ④

Kontakt zu Ärzt:innen

b3

3 ... zum Personal betreffend Therapie und Seniorenbetreuung ① ② ③

Kontakt zu anderen Personal (Therapie, Seniorenbetreuung)

Vom Fragebogen zur Datenmatrix – Datenlabeling

Variablenname – Variablenlabel

Werte/Codes – Wertelabels

a1

Geschlecht



1 männlich
2 weiblich
99 keine Antwort

a2

Alter



15 bis 99
999 keine Angabe

b1

Kontakt zu Pflegepersonal



① stimme voll zu
②
③
④
⑤
⑥ stimme nicht zu
-1 keine Antwort

Den Fragebogen übersichtlich in Teile gliedern, in der Datenmatrix mit Buchstaben kennzeichnen – erleichtert die Dateneingabe und Auswertung !

Codierung des Fragebogens

1. Ich habe mich in meinem Praktikum wohl gefühlt

V1a ja 1 eher ja 2 eher nein 3 nein 4

v1b Bitte geben Sie eine Erklärung dazu.

.....
.....

2. Zu Beginn meines Praktikums erfolgte ein Einführungsgespräch.

V2a ja 1 nein 2

Das Einführungsgespräch war informativ.

V2b ja 1 eher ja 2 eher nein 3 nein 4

V2c Es gab nützliche schriftliche Einführungsunterlagen.

ja 1 nein 2

V3a 3. Es erfolgte ein Zwischengespräch.

ja 1 nein 2

v3b Das Zwischengespräch beinhaltete ein aufschlussreiches Feedback meiner Leistungen.

ja 1 eher ja 2 eher nein 3 nein 4

Datenmatrix – allgemeine Schreibweise

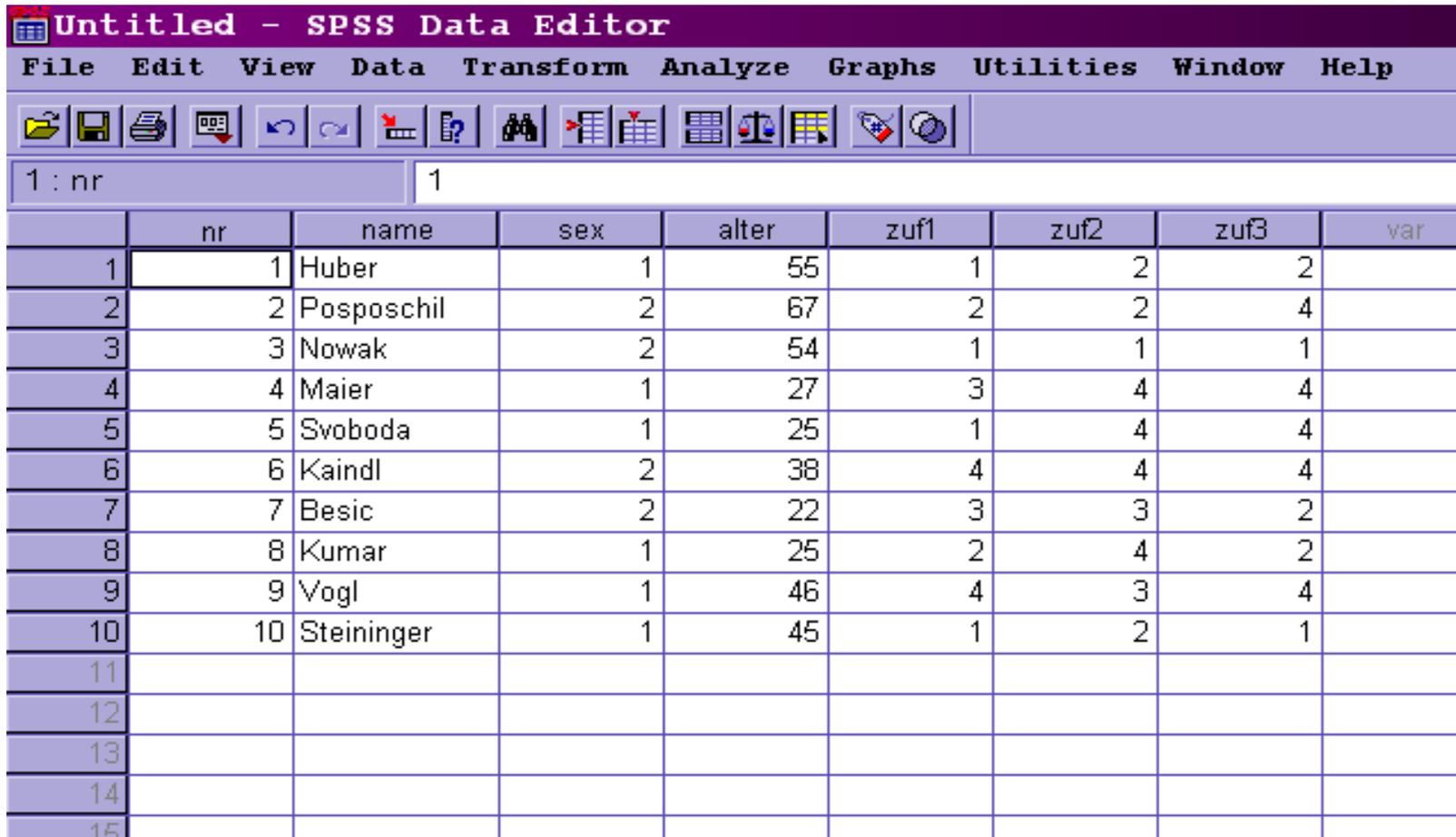
		Variablen					
		var1	var2	var3	.	.	var _k
Individuelle Beobachtungen (Personen) ↓	1	X₁₁	X₂₁	X₃₁	.	X_{k1}	X_{k1}
	2	X₁₂	X₂₂	X₃₂	.	X_{k2}	X_{k2}
	3	X₁₃	X₂₃	X₃₃	.	X_{k3}	X_{k3}
	4	X₁₄	X₂₄	X₃₄	.	X_{k4}	X_{k4}

	i	X_{1i}	X_{2i}	X_{3i}	.	X_{ki}	X_{ki}

Datenmatrix – Antworten als Zahlen codiert

		Variablen		
		sex	famst	alter
Individuelle Beobachtungen (Personen) ↓	1	1	2	35
	2	2	4	67
	3	2	3	33
	4	1	1	30

Datenansicht



1 : nr 1

	nr	name	sex	alter	zuf1	zuf2	zuf3	var
1	1	Huber	1	55	1	2	2	
2	2	Posposchil	2	67	2	2	4	
3	3	Nowak	2	54	1	1	1	
4	4	Maier	1	27	3	4	4	
5	5	Svoboda	1	25	1	4	4	
6	6	Kaindl	2	38	4	4	4	
7	7	Besic	2	22	3	3	2	
8	8	Kumar	1	25	2	4	2	
9	9	Vogl	1	46	4	3	4	
10	10	Steininger	1	45	1	2	1	
11								
12								
13								
14								
15								

Variablenansicht

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	nr	Numeric	11	0		None	None	8	Right	Scale
2	name	String	10	0		None	None	10	Left	Nominal
3	sex	Numeric	11	0		None	None	8	Right	Scale
4	alter	Numeric	11	0		None	None	8	Right	Scale
5	zuf1	Numeric	11	0		None	None	8	Right	Scale
6	zuf2	Numeric	11	0		None	None	8	Right	Scale
7	zuf3	Numeric	11	0		None	None	8	Right	Scale
8										
9										

	Name	Type	Width	Dec	Label	Values	Missing	Colu	Align	Measure
1	nr	Numeric	3	0	Fragebogenlaufnummer	None	None	8	Right	Scale
2	name	String	25	0	Name	None	None	10	Left	Nominal
3	sex	Numeric	3	0	Geschlecht	{1, männlich}...	None	8	Right	Nominal
4	alter	Numeric	3	0	Alter in Jahren	None	None	8	Right	Ordinal
5	zuf1	Numeric	3	0	Zufrieden mit der BeraterIn	{1, sehr zufrieden}...	99	8	Right	Ordinal
6	zuf2	Numeric	3	0	Zufrieden mit den Öffnungszeiten	{1, sehr zufrieden}...	99	8	Right	Ordinal
7	zuf3	Numeric	3	0	Zufrieden mit der Beratungsqualität	{1, sehr zufrieden}...	99	8	Right	Ordinal
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										



Value Labels ? X

Value Labels:

Value:

Value Label:

1 = "sehr zufrieden"
 2 = "eher zufrieden"
 3 = "eher nicht zufrieden"
 4 = "gar nicht zufrieden"
 99 = "keine Antwort"

Datenmatrix – Studi-FB

	sem	nr	S1	S2	S3	S4	S5	S5_txt	S6	S6_di	S7	S8	S9	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11
1	1	1	2	38	5	japanisch	2	Japan	1	1	5	3	2	6	6	6	4	6	3	1	5	7	4	5
2	1	2	2	24	5	deutsch	1		1	1	7	2	6	5	4	6	2	5	4	1	5	4	5	3
3	1	3	1	22	5	deutsch	1		1	1	5	2	3	5	7	7	4	5	6	1	7	2	6	0
4	1	4	2	22	2	deutsch	1		2	2	3	2	2	4	4	5	6	4	3	2	5	4	5	4
5	1	5	2	22	5	deutsch	2	Deutschla	1	1	10	1	2	5	5	7	6	5	4	1	5	4	4	4
6	1	6	2	20	4	österreich	1		1	1	9	2	2	5	5	6	6	6	3	4	7	6	5	5
7	1	7	2	20	4	deutsch	1		1	1	5	2	1	6	4	6	6	5	4	2	3	5	3	3
8	1	8	2	20	4	deutsch	1		1	1	5	2	3	6	6	7	5	5	4	2	4	5	2	6
9	1	9	1	22	5	deutsch	1		1	1	10	2	6	5	4	7	3	6	4	1	6	7	6	6
10	1	10	2	20	5	deutsch	1		1	1	5	1	3	4	4	5	5	4	3	2	3	4	4	3
11	1	11	2	27	4	deutsch	1		2	2	5	2	2	4	5	4	5	5	5	3	5	5	5	5
12	1	12	2	21	4	deutsch	1		2	2	5	2	2	5	5	4	4	4	3	1	3	5	5	2
13	1	13	2	27	5	deutsch	1		2	2	8	2	2	6	6	6	5	5	4	3	4	5	5	4
14	1	14	1	35	5	deutsch	1		2	2	10	2	3	5	5	5	5	6	5	2	4	4	4	2
15	1	15	2	25	5	deutsch	1		1	1	7	2	2	7	7	6	5	5	6	2	7	6	4	3
16	1	16	2	24	2	deutsch	1		2	2	5	1	2	5	7	7	7	6	6	1	5	6	5	5
17	1	17	2	21	5	deutsch	1		1	1	5	2	2	4	5	7	7	6	6	3	6	6	6	5
18	1	18	1	25	5	deutsch	1		2	2	8	1	2	5	5	5	4	4	4	3	3	3	5	3
19	1	19	2	22	3	deutsch	2	Italien	1	1	7	2	3	5	6	7	4	6	5	3	5	5	6	4
20	1	20	2	24	2	deutsch	1		1	1	4	2	6	5	4	5	6	6	3	2	3	4	4	4
21	1	21	2	26	4	deutsch	2	Deutschla	2	2	5	1	2	5	5	6	6	6	6	2	6	6	6	6
22	1	22	2	22	3	deutsch	1		1	1	8	2	3	5	5	5	5	5	5	4	6	6	5	5
23	1	23	2	34	5	deutsch	1		4	2	5	2	4	5	5	6	0	5	5	2	5	6	6	5
24	1	24	1	24	0	deutsch	1		2	2	8	1	2	4	4	4	4	4	4	4	4	4	4	4
25	1	25	1	26	5	österreich	1		2	2	1	2	2	6	4	3	3	5	3	4	2	3	3	3
26	1	26	1	32	4	deutsch	1		1	1	5	2	3	7	5	7	4	6	6	2	6	6	3	5
27	1	27	2	24	2	deutsch	1		1	1	5	2	2	6	7	6	7	7	7	1	7	7	7	6
28	1	28	2	25	2	deutsch	1		3	2	6	2	2	5	6	6	6	6	6	7	6	6	5	5
29	1	29	1	24	4	deutsch	1		1	1	4	1	6	4	3	4	6	5	3	1	3	3	3	2

Datenmatrix – Studie Kopfschmerzambulanz

Menz, Florian (2013): Zum Vergleich von ärztlichen Konsultationen zu Kopfschmerzen bei gedolmetschten und nicht gedolmetschten Gesprächen. In Florian Menz (Ed.), Migration und medizinische Kommunikation (pp. 311-352). Wien: v&r Vienna University Press.

1 : Inr		1																	Sichtbar: 49 von 49 Varia	
	dauer_min	dauer_sec	Partituranzahl	partanz_kat	Wortanzahl_P	Wortanzahl_A	wort_P	wort_A	Wortanzahl_LD	Wortanzahl_P_D	wort_LD	wort_P_D	dolm	dolm_kat	summ_P_A	summ_P_A_kat	summ_P_A_D			
1	29	59	99	99	99	99	99	99	77	77	3	3	2	3	99	99	99			
2	21	31	99	99	99	99	99	99	77	77	3	3	2	3	99	99	99			
3	24	44	266	2	724	1502	1	1	787	77	1	3	1	2	2226	1	3013			
4	40	0	422	3	785	3506	2	2	1123	77	2	3	1	2	4291	2	5414			
5	28	11	380	3	2377	2459	2	2	77	77	3	3	2	3	4836	2	77			
6	56	15	545	3	99	99	99	99	77	77	3	3	2	3	99	99	99			
7	41	35	468	3	2412	3996	2	2	77	77	3	3	2	3	6408	2	77			
8	34	56	365	3	99	99	99	99	77	77	3	3	2	3	99	99	99			
9	25	54	99	99	99	99	99	99	77	77	3	3	2	3	99	99	99			
10	42	27	410	3	2711	2914	2	2	77	77	3	3	2	3	5625	2	77			
11	31	54	486	3	2615	3744	2	2	77	77	3	3	2	3	6359	2	77			
12	39	59	498	3	1048	2257	2	2	77	3238	3	2	1	1	3305	2	6543			
13	18	20	172	1	1524	1446	2	1	77	77	3	3	2	3	2970	1	77			
14	12	6	99	99	99	99	99	99	77	77	3	3	2	3	99	99	99			
15	36	20	99	99	99	99	99	99	77	77	3	3	2	3	99	99	99			
16	26	8	312	2	99	99	99	99	77	77	3	3	2	3	99	99	99			
17	22	3	251	2	99	99	99	99	77	77	3	3	2	3	99	99	99			
18	24	24	225	1	480	1111	99	99	77	77	3	3	2	3	99	99	99			
19	31	53	341	2	1846	2908	2	2	77	77	3	3	2	3	4754	2	77			
20	38	36	497	3	99	99	99	99	77	77	3	3	2	3	99	99	99			
21	6	18	49	1	99	99	99	99	77	77	3	3	2	3	99	99	99			
22	14	21	99	99	99	99	99	99	77	77	3	3	2	3	99	99	99			
23	47	32	487	3	135	3610	1	2	3296	77	2	3	1	2	3745	2	7041			
24	11	38	125	1	203	897	1	1	572	77	1	3	1	2	1100	1	1672			
25	16	50	147	1	381	675	1	1	784	77	1	3	1	2	1056	1	1840			
26	24	0	263	2	1496	2028	2	2	77	77	3	3	2	3	3524	2	77			
27	64	1	99	99	99	99	99	99	77	77	3	3	2	3	99	99	99			
28	27	37	316	2	587	1676	1	1	141	1837	1	2	1	1	2263	1	4241			
29	5	56	60	1	99	99	99	99	77	77	3	3	2	3	99	99	99			

Datenquelle: FWF-Projekt "Pain representation", Leitung Florian Menz

Labeling

Name	Kurzbezeichnung der Variable (erstes Zeichen ein Buchstabe)
Typ	Numerisch oder String (Text/Zeichenfolge)
Spaltenformat (1-40)	Anzahl der Stellen vor dem Komma
Dezimalstellen (0-16)	Anzahl der Dezimalstellen
Variablenlabel	Beschreibung der Variable
Wertelabels	Bezeichnung der einzelnen Ausprägungen (Werte)
Fehlende Werte	Definieren von fehlenden Werten
Spalten (1 bis 255)	Breite der Variable im Datenfenster
Ausrichtung	Links, Rechts, Mitte
Messniveau	Nominal, Ordinal, Metrisch

Dateneingabe

→ Jeder FB erhält eine **Laufnummer**: wichtig für die Nachvollziehbarkeit und zur Kontrolle bei Eingabefehlern

→ Jede Variable benennen: möglichst kurz: **Variablenname**
Variablenlabel : Beschreibung der Variable - kann länger sein

→ **Codeplan** erstellen:
- jeder Antwortmöglichkeit wird eine Zahl zugeordnet
- Optimal: bei der FB-Erstellung die Fragen nummerieren und die Codes (**Werte und Wertelabels**) darstellen

→ Während der Eingabe: zwischendurch immer wieder speichern
→ Nach der Dateneingabe: **Datenkontrolle** mittels Häufigkeitstabellen
Falsche Einträge nach Möglichkeit nachkontrollieren oder löschen

→ Regeln zum **Umgang mit fehlenden Eintragungen** aufstellen:
Entweder leer lassen oder festgelegte Codes eintragen (z.B. 99)

SPSS
Statistical **P**rogram for **S**ocial **S**cientists

OUTPUT

Dokumentation
der Ergebnisse

Tabellen, Text,
Grafiken, Diagramme

Kopieren
ins **Word** mit Strg+k
(als Objekt)
ins **Excel** mit Strg+c
(als bearbeitbare Tabelle)

***.spv**

DATENMATRIX

Datenansicht
Spalten: Variablen
Zeilen: Beobachtungen

Variablenansicht
Auflistung der Variablen in
Reihenfolge des Datensatzes
(Tabellensheet)

Daten eingeben
Variablen und Werte labeln
Variablen transformieren
(berechnen, umkodieren ...)

***.sav**

SYNTAX

Dokumentation der
Auswertungsschritte

Speichern der Befehle
(mit „paste/Einf.“)

Befehle nachvollziehen
Befehlsabfolgen
wiederholen

***.sps**

Häufigkeitstabelle

Kategorienbildung

Echtdatensatz: Gespräche in der Kopfschmerzambulanz

n = 75 (Menz 2010)

Analysieren → deskriptive Statistiken → Häufigkeiten ... unter_art

Häufigkeitstabelle

unter_art Art erste externe Unterbrechung

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 InstitutsangehörigeR betritt das Zimmer	22	29,3	36,1	36,1
	2 Pager mit Rückruf	7	9,3	11,5	47,5
	3 Pager ohne Rückruf	1	1,3	1,6	49,2
	4 Telefonanruf	9	12,0	14,8	63,9
	5 keine Unterbrechung	22	29,3	36,1	100,0
	Gesamt	61	81,3	100,0	
Fehlend	99 keine Angabe	14	18,7		
Gesamt		75	100,0		

Berechnen von Prozentwerten bei einer Häufigkeitstabelle

Gültige Fälle $g = 61$

Fehlende Fälle $f = 14$

Gesamt $n = g + f = 75$

Häufigkeit Häufigkeit des Auftretens einzelner Variablenwerte:
 n_1 (IA betritt Zimmer) = 22; n_2 (Pager mit Rückruf) = 7 Fälle usw.

% Prozentverteilung der Werte über alle Fälle:
Anteil der Gespräche wo IA Zimmer betritt:

$$p_1 = \frac{n_1}{n} 100 = \frac{22}{75} 100 = 29,3\%$$

Gültige % Prozentverteilung der Werte über alle Fälle mit gültigen Werten:
Anteil der Gespräche wo IA Zimmer betritt an allen Gespr. mit gültigen Werten:

$$v_1 = \frac{n_1}{g} 100 = \frac{22}{61} 100 = 36,1\%$$

Kumulierte % Kumulierte Prozentverteilung über alle Fälle mit gültigen Werten:
Anteil der Gespräche wo IA Zimmer betritt:

$$k_1 = \frac{n_1}{g} 100 = \frac{22}{61} 100 = 36,1\%$$

Anteil der Gespräche wo IA Zimmer betritt + mit Pager mit Rückruf:

$$k_2 = \frac{(n_1 + n_2)}{g} 100 = \frac{(22 + 7)}{61} 100 = 47,5\%$$

Anteil aller Gespräche mit Unterbrechungen (IA + Pager m/o + Telefon):

$$k_3 = \frac{(n_1 + n_2 + n_3 + n_4)}{g} 100 = \frac{(22 + 7 + 1 + 9)}{61} 100 = 63,9\%$$

Mehrfachantworttabelle –

Belastende Ereignisse in der Kindheit

	Anzahl	Anzahl als Spalten (%)	Antworten	Antworten als Spalten (%) (Basis: Anzahl)	Anzahl als Spalten (%) (Basis: Antworten)
b1 separation or divorce of parents	22	68,8%	22	68,8%	11,9%
b2 death of mother or father	10	31,3%	10	31,3%	5,4%
b3 alcohol, medicament or drug abuse of father or mother	16	50,0%	16	50,0%	8,6%
b4 grow up in a foster family	5	15,6%	5	15,6%	2,7%
b5 physical violence or threats from family members, relatives or known persons	15	46,9%	15	46,9%	8,1%
b6 sexual harassments or sexual abuse from family members, relatives or known persons	14	43,8%	14	43,8%	7,6%
b7 runaway from home as child or adolescent because of conflicts	19	59,4%	19	59,4%	10,3%
b8 death of a close girlfriend or boy friend	24	75,0%	24	75,0%	13,0%
b9 serious sickness of a close person	17	53,1%	17	53,1%	9,2%
b10 own serious sickness	24	75,0%	24	75,0%	13,0%
b11 attempted suicide	18	56,3%	18	56,3%	9,7%
b0 none	1	3,1%	1	3,1%	0,5%
Gesamt	32	100,0%	185	578,1%	100,0%

32 Befragte haben 185 belastende Ereignisse in der Kindheit erlebt, das sind durchschnittlich 5,781 Ereignisse pro Befragten.

75% von allen Befragten haben eine/n enge/n FreundIn durch Tod verloren.

13% aller belastenden Ereignisse betrafen den Verlust eines/r engen FreundIn.

Mehrfachantworttabelle – Interessensgebiete bei Nachrichten

attention to new s	Anzahl	Anzahl der Spalten (%)
about_Social issues	1310	67,3%
about_National politics	1087	55,8%
about_The economy	1209	62,1%
about_Foreign policy	759	39,0%
about_International relationship	831	42,7%
about_Local politics	901	46,3%
about_Sport	1017	52,2%
about_Culture	1041	53,5%
about_The environment	965	49,6%
about_EU	721	37,0%
esamt	1947	100,0%

Anzahl_attention to news

	Häufigkeit	Gültige Prozepte	Kumulierte Prozepte
0	53	2,7	2,7
1	181	9,1	11,7
2	260	13,0	24,7
3	282	14,1	38,8
4	248	12,4	51,2
5	167	8,4	59,6
6	173	8,7	68,2
7	173	8,7	76,9
8	149	7,5	84,3
9	114	5,7	90,0
10	200	10,0	100,0
Gesamt	2000	100,0	

67% der Befragten haben Interesse an Sozialen Themen.

- 2,7% der Befragten haben an keinem der Themen Interesse.
- 10% der Befragten haben an allen 10 Themen Interesse.

Mehrfachantworttabelle – regelmäßig gelesene Zeitungen

\$d1r Regelmäßig gelesene Zeitungen	Anzahl	Anzahl als Spalten (%)
d1r_keine keine Zeitung regelmäßig	1365	30,3%
d1r_1 Kronen Zeitung	1278	28,4%
d1r_2 Kleine Zeitung	601	13,3%
d1r_8 Kurier	426	9,5%
d1r_12 Salzburger Nachrichten	206	4,6%
d1r_3 OÖ Nachrichten	193	4,3%
d1r_4 Tiroler Tageszeitung	193	4,3%
d1r_5 Die Presse	186	4,1%
d1r_15 Der Standard	169	3,8%
d1r_16 Vorarlberger Nachrichten	162	3,6%
d1r_11 Österreich	159	
d1r_9 TT Kompakt	44	
d1r_14 Wirtschaftsblatt	34	
d1r_17 Heute	22	
d1r_6 Neues Volksblatt	16	
d1r_10 Wiener Zeitung	16	
d1r_7 Die ganze Woche	14	
d1r_13 Medianet	14	
d1r_18 Sportwoche	10	
d1r_20 Falter	1	
d1r_19 Furche	0	
Gesamt	4504	

28% der Befragten lesen
regelmäßig die Kronen Zeitung.

58% der Befragten lesen
eine Zeitung regelmäßig.

**anzahl_regelm Anzahl regelmäßig gelesener
Tageszeitungen**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0	1365	30,3	30,3	30,3
	1	2625	58,3	58,3	88,6
	2	442	9,8	9,8	98,4
	3	58	1,3	1,3	99,7
	4	10	,2	,2	99,9
	5	3	,1	,1	100,0
	6	1	,0	,0	100,0
Gesamt		4504	100,0	100,0	

Checkliste: Interpretieren einer Häufigkeitstabelle

- Erwähnen von Fallzahl und Anteil der fehlenden Antworten
- Immer die gültigen Prozente interpretieren
(Ausnahme: die fehlenden Werte haben besondere Relevanz)
- Interpretation bedeutet nicht die Auflistung von Prozenten!
- Interpretation bedeutet die inhaltliche Zusammenfassung der Ergebnisse!
- Gleichzeitig ist aber Interpretation ohne Werte nicht komplett – Werte alleine reichen aber auch nicht aus!

**Lage- und Streuungsmaße:
Modus, Median, Quartile
Mittelwert, Varianz
Verteilungsformen**

Wortanzahl_A Wortanzahl ÄrztIn

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	480	1	1,3	2,7	2,7
	493	1	1,3	2,7	5,4
	542	1	1,3	2,7	8,1
	675	1	1,3	2,7	10,8
	748	1	1,3	2,7	13,5
	752	1	1,3	2,7	16,2
	897	1	1,3	2,7	18,9
	1101	1	1,3	2,7	21,6
	1256	1	1,3	2,7	24,3
	1310	1	1,3	2,7	27,0
	1446	1	1,3	2,7	29,7
	1461	1	1,3	2,7	32,4
	1502	1	1,3	2,7	35,1
	1674	1	1,3	2,7	37,8
	1676	1	1,3	2,7	40,5
	1752	1	1,3	2,7	43,2
	1762	1	1,3	2,7	45,9
	1886	1	1,3	2,7	48,6
	1930	1	1,3	2,7	51,4
	2028	1	1,3	2,7	54,1
	2064	1	1,3	2,7	56,8
	2145	2	2,7	5,4	62,2
	2257	1	1,3	2,7	64,9
	2276	1	1,3	2,7	67,6
	2459	1	1,3	2,7	70,3
	2579	1	1,3	2,7	73,0
	2727	1	1,3	2,7	75,7
	2908	1	1,3	2,7	78,4
	2914	1	1,3	2,7	81,1
	3001	1	1,3	2,7	83,8
	3434	1	1,3	2,7	86,5
	3506	1	1,3	2,7	89,2
	3610	1	1,3	2,7	91,9
	3744	1	1,3	2,7	94,6
	3996	1	1,3	2,7	97,3
	4109	1	1,3	2,7	100,0
	Gesamt	37	49,3	100,0	
Fehlend	99 keine Angabe	38	50,7		
Gesamt		75	100,0		

1. Quartil

Median

3. Quartil

MW?

Analysieren → deskriptive Statistiken → Häufigkeiten ... wort_A_in200

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 400-600	3	4,0	8,1	8,1
	2 601-800	3	4,0	8,1	16,2
	3 801-1000	1	1,3	2,7	18,9
	4 1001-1200	1	1,3	2,7	21,6
	5 1201-1400	2	2,7	5,4	27,0
	6 1401-1600	3	4,0	8,1	35,1
	7 1601-1800	4	5,3	10,8	45,9
	8 1801-2000	2	2,7	5,4	51,4
	9 2001-2200	4	5,3	10,8	62,2
	10 2201-2400	2	2,7	5,4	67,6
	11 2401-2600	2	2,7	5,4	73,0
	12 2601-2800	1	1,3	2,7	75,7
	13 2801-3000	2	2,7	5,4	81,1
	14 3001-3200	1	1,3	2,7	83,8
	15 3201-3400	1	1,3	2,7	86,5
	16 3401-3600	1	1,3	2,7	89,2
	17 3601-3800	2	2,7	5,4	94,6
	18 3801-4000	1	1,3	2,7	97,3
	19 4001-4200	1	1,3	2,7	100,0
Gesamt		37	49,3	100,0	
Fehlend	99 keine Angabe	38	50,7		
Gesamt		75	100,0		

Modi →
Modi →

← 1. Quartil
← Median
← 3. Quartil

Statistische Kennzahlen: Lagemaße

Beschreiben die Lage des Gipfels der Verteilung

Modus = der am häufigsten genannte/beobachtete Wert

Median = Wert, der die Stichprobe in die Hälfte teilt:

50% der Fälle liegen **unter** dem Median,
50% der Fälle liegen **über** dem Median

Quartile = teilen die Stichprobe in Viertel:

25% der Fälle liegen bis zum 1. Quartil

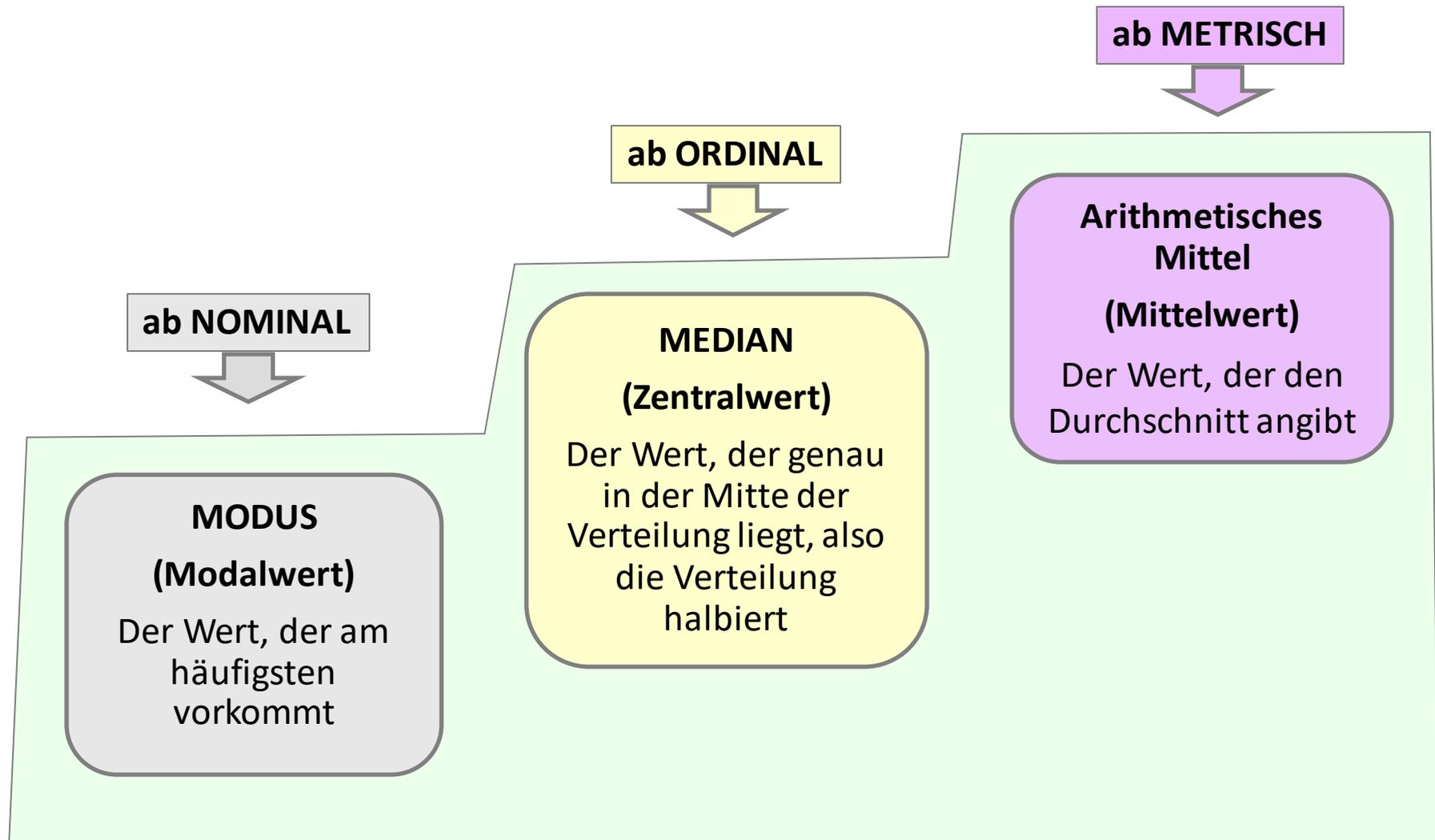
50% der Fälle liegen bis zum 2. Quartil (= Median)

75% der Fälle liegen bis zum 3. Quartil

Das restliche Viertel der Fälle liegt schließlich über dem 3. Quartil.

Statistische Kennzahlen: Lagemaße

Beschreiben die Lage des Gipfels einer Verteilung



Mittelwert = Durchschnitt

Jene Ausprägung, die alle hätten, wenn alle „gleich viel“ von dem Merkmal hätten.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Summe
aller Messwerte

Anzahl
der Messwerte

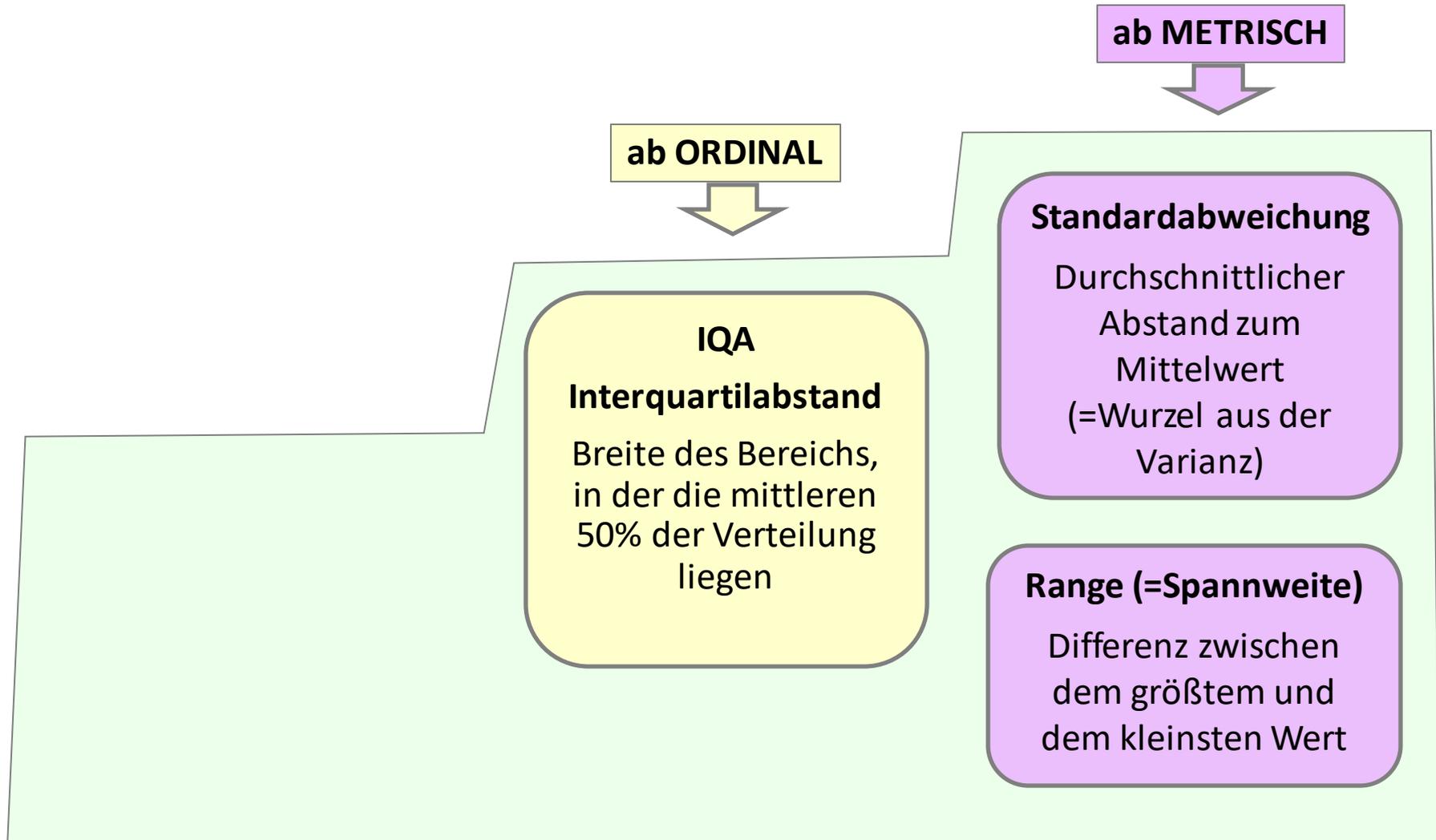
$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + x_4 \dots \dots \dots + x_n$$

z.B. **Durchschnittliches Monats-Einkommen:**

Summe von allen monatlichen Einkommen – „gerecht“ aufteilen (pro Kopf)

Statistische Kennzahlen: Streuungsmaße

Beschreiben die Breite einer Verteilung



Varianz / Standardabweichung² = Streuung um den Mittelwert

Wie weit streuen die Messwerte um den Mittelwert?

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

Summe aller quadrierten
Abweichungen
zum Mittelwert

Anzahl
der Messwerte

Varianz

Standardabweichung

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

$$\sum_{i=1}^n (x_i - \bar{x})^2 =$$

$$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$$

Statistische Kennzahlen: Streuungsmaße

Beschreiben die Breite der Verteilung

Spannweite = Abstand zwischen kleinstem und größtem Wert

Quartilabstände = Abstände zwischen den einzelnen Quartilen
Interquartilabstand = Abstand zwischen 1. und 3. Quartil:
in diesem Bereich liegen die mittleren 50% der Verteilung

Standardabweichung = durchschnittliche Streuung um den Mittelwert
Varianz = quadrierte durchschnittliche Streuung um den Mittelwert

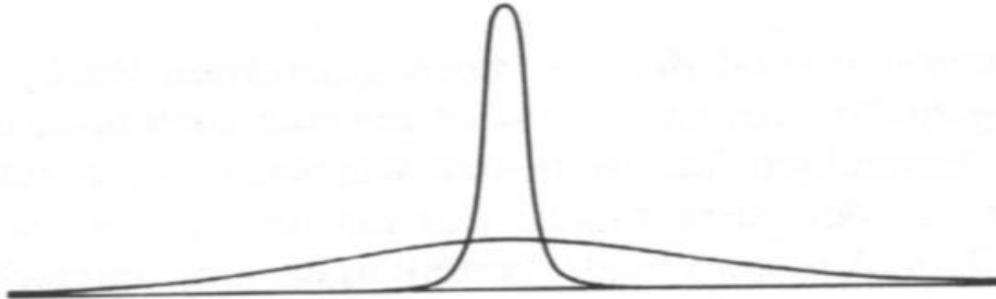
$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Varianz

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

Standardabweichung

**Problem: Wenn ich nur den Mittelwert weiß,
weiß ich noch nichts über die Breite der Streuung!**



Gedankenexperiment:

Was kommt heraus, wenn von allen Messwerten die Abstände zum Mittelwert berechnet und summiert werden?

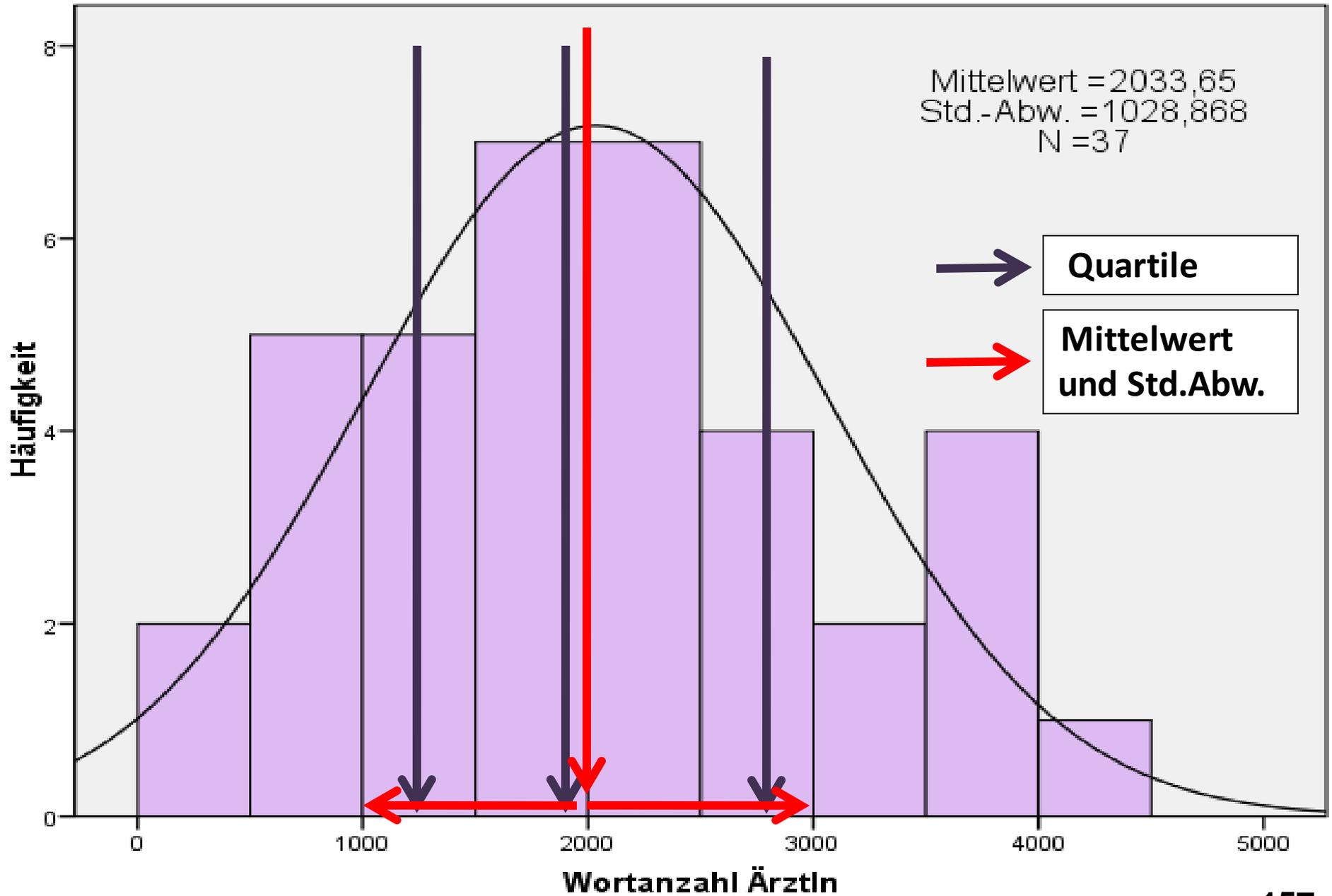
$$\sum_{i=1}^n (x_i - \bar{x}) =$$

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + (x_4 - \bar{x}) \dots \dots \dots + (x_n - \bar{x}) \quad = ???$$

Statistiken

Wortanzahl_A Wortanzahl Ärztl

N	Gültig	37
	Fehlend	38
Mittelwert		2033,65
Median		1930,00
Modus		2145
Standardabweichung		1028,868
Varianz		1058570
Spannweite		3629
Minimum		480
Maximum		4109
Summe		75245
Perzentile	25	1283,00
	33,33333333	1488,33
	50	1930,00
	66,66666667	2337,00
	75	2817,50



Interpretation der Verteilung der Wortanzahl der Ärzt:innen

Die Verteilung ist leicht **linksgipfelig** (= linkssteil, rechtsschief)

– der Gipfel liegt stärker bei den niedrigeren Ausprägungen.

In 50% der Gespräche haben die Ärzt:innen bis zu 1900 Wörter gesprochen.

In den anderen 50% der Gespräche haben sie mehr als 1900 Wörter gesprochen.

Durchschnittlich wurden von den Ärzt:innen 2000 Wörter in einem Gespräch gesprochen.

Der **Mittelwert** ist nur geringfügig höher als der Median, was dafür spricht, dass es ein paar Ausreißer nach oben gibt (Gespräche, wo die Ärzt:innen sehr viel sprechen).

Dies lässt auch darauf schließen, dass die Verteilung leicht schief ist (fast symmetrisch), und zwar leicht linksgipfelig (oder = linkssteil, rechtsschief).

Die **Streuung** beträgt 1029 Wörter um den Mittelwert, was bei einer Spannweite von 3629 Wörtern eine relativ breite Streuung der Verteilung darstellt.

Checkliste: Interpretieren statistischer Kennzahlen

- Welche Daten/ welche Variable wird dargestellt?
Konkret angeben, um was es sich handelt.
- Erwähnen von Fallzahl und Anteil der fehlenden Antworten
- Interpretation bedeutet nicht die Auflistung von Werten!
- Interpretation bedeutet die jeweilige Kennzahl auf die konkreten Daten zu beziehen!
- Gleichzeitig ist aber Interpretation ohne Werte nicht komplett – Werte alleine reichen aber auch nicht aus!

Falsch:

Der Median beträgt 1930.

Richtig: Bei 50% der untersuchten 37 Gespräche in der Kopfschmerzambulanz haben die Ärzt:innen bis zu 1930 Wörter gesprochen, bei der anderen Hälfte mehr als 1930 Wörter.

Statistiken

		Wortanzahl_A Wortanzahl ÄrztIn	Wortanzahl_P Wortanzahl PatientIn
N	Gültig	37	37
	Fehlend	38	38
Mittelwert		2033,65	1054,68
Median		1930,00	758,00
Modus		2145	135 ^a
Standardabweichung		1028,868	824,621
Varianz		1058570,234	679999,003
Schiefe		,349	1,420
Standardfehler der Schiefe		,388	,388
Kurtosis		-,697	1,633
Standardfehler der Kurtosis		,759	,759
Minimum		480	135
Maximum		4109	3641
Summe		75245	39023
Perzentile	25	1283,00	479,50
	50	1930,00	758,00
	75	2817,50	1414,00

Wortanzahl von ÄrztInnen und PatientInnen im Vergleich

Ärzt:innen sprechen durchschnittlich 2033 Wörter, Patient:innen durchschnittlich 1054 Wörter.

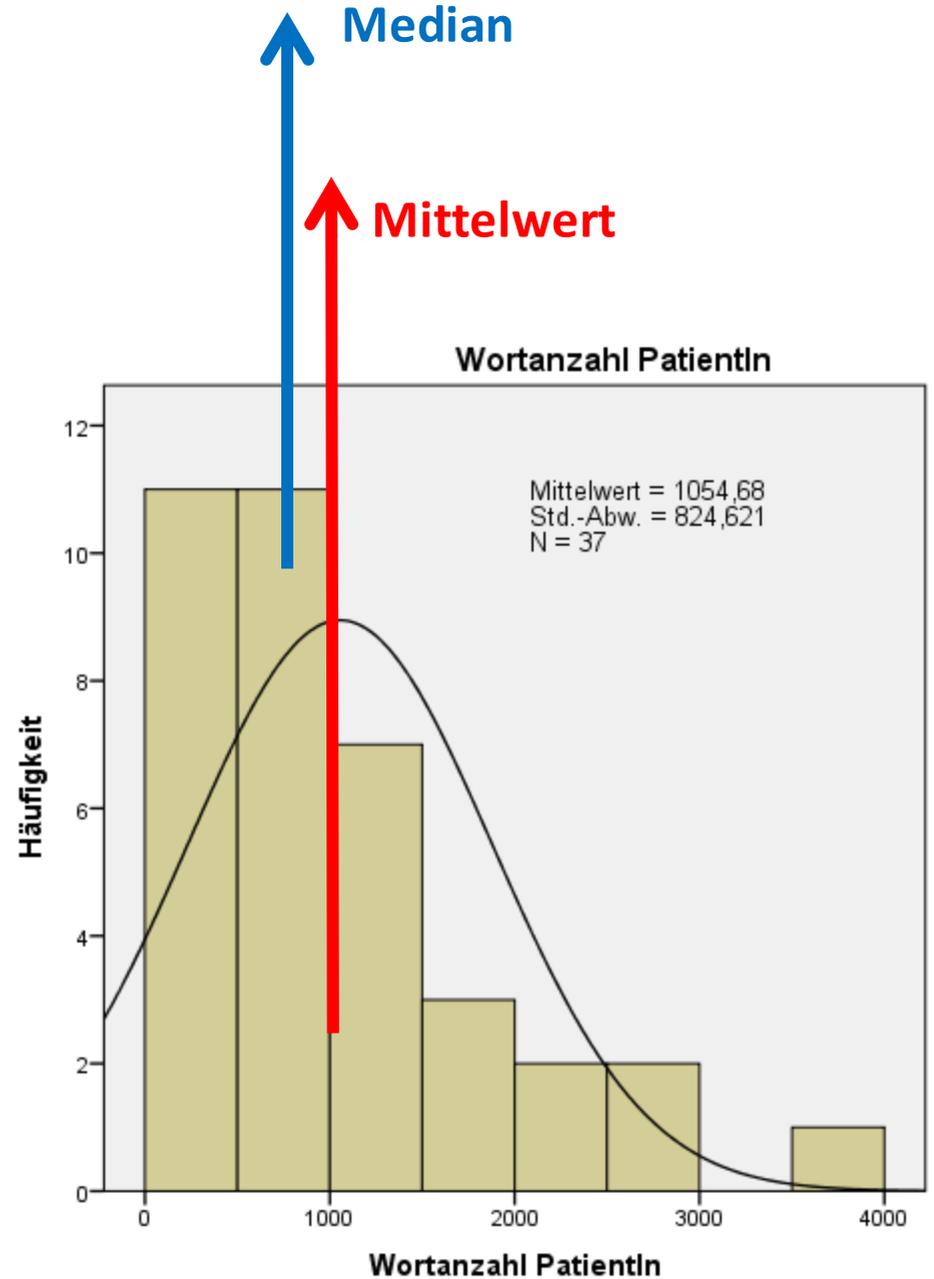
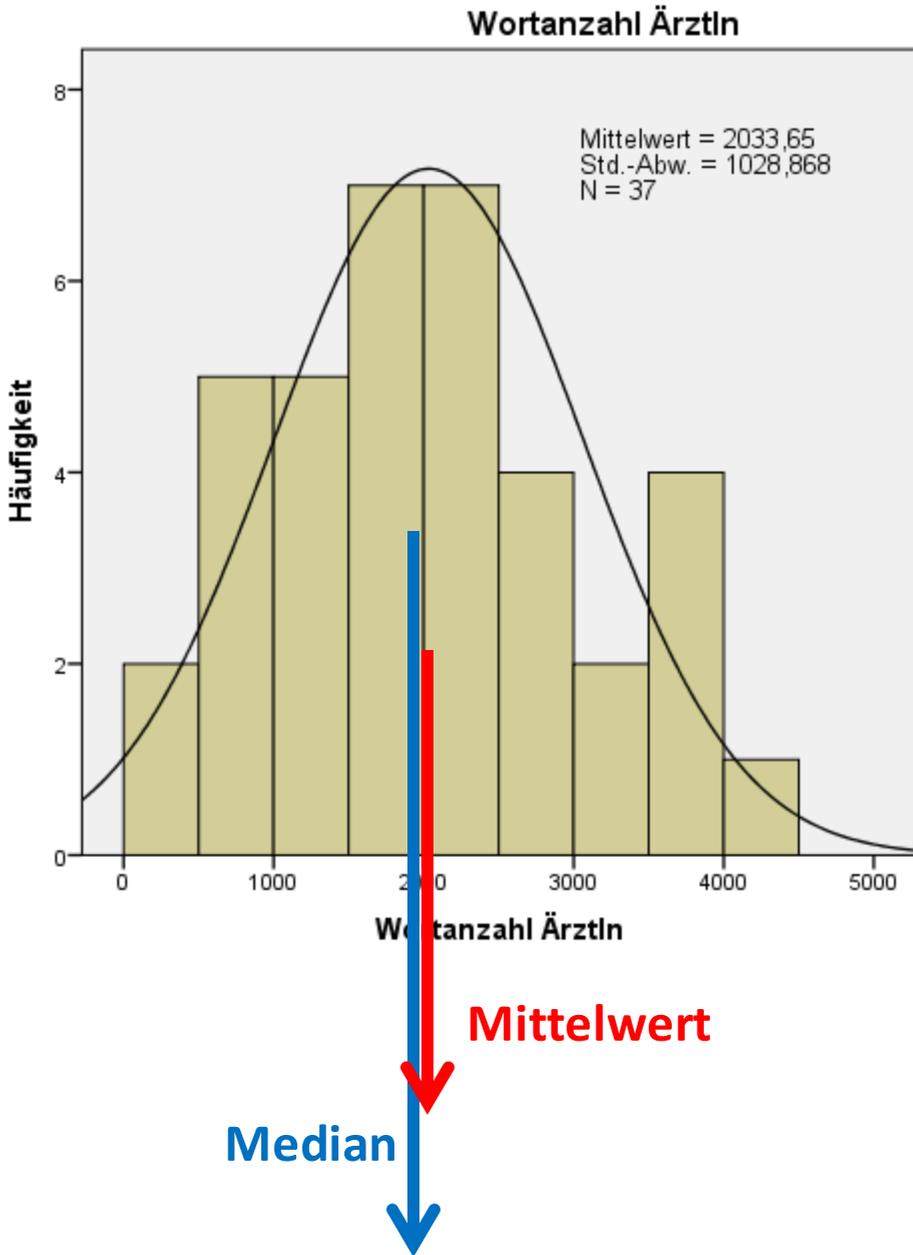
Bei den Ärzt:innen sind Mittelwert und Median fast gleich, bei den Patient:innen ist der Median deutlich niedriger. (linksgipfelig?)

Die Streuung um den Mittelwert ist bei den Ärzt:innen größer (1029 Wörter) – mehr Bandbreite in der Wortanzahl.

Verteilung Ärzt:innen: symmetrisch (nahe 0)
Verteilung Patient:innen: linksgipfelig (>0)

Verteilung Ärzt:innen: wie NV (nahe 0)
Verteilung Patient:innen: steiler als NV (>0)

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.



Datenniveaus und statistische Kennzahlen

Maßeinheit mit gleichen Abständen

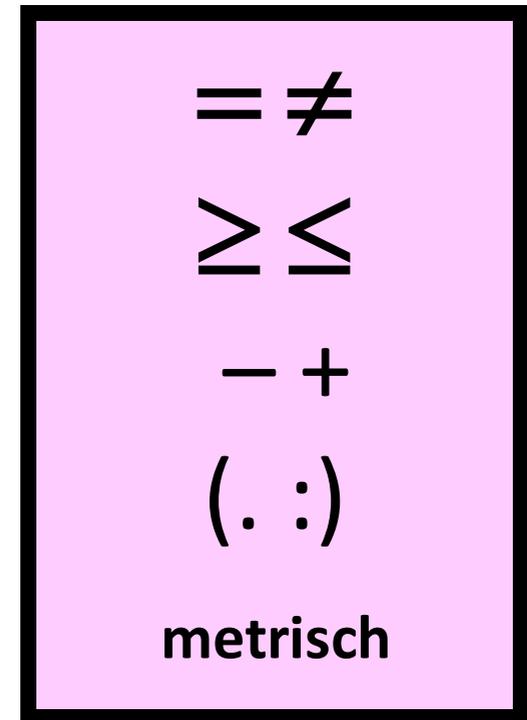
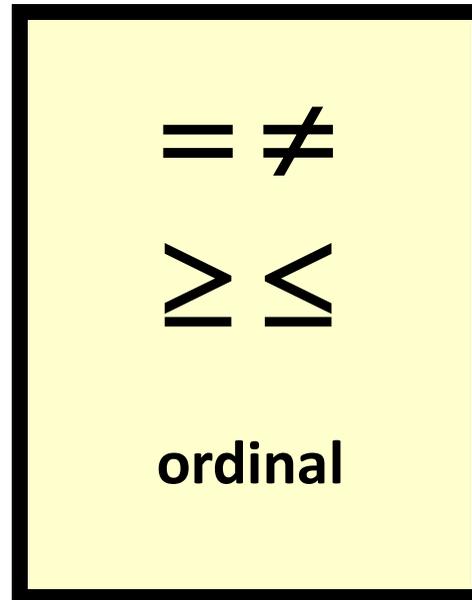
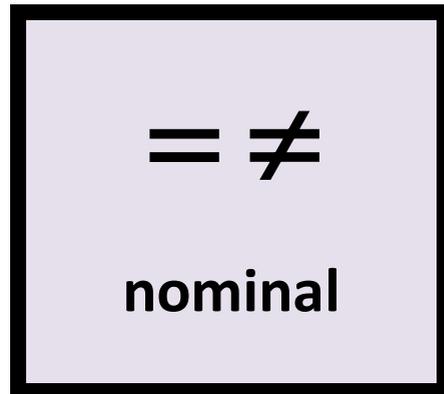
Mittelwert
Varianz/Standardabweichung
Spannweite
Schiefe / Wölbung

Reihenfolge

Median, Quartile
Interquartilabstand

Gruppen

Modus



Übersicht über Lage- und Streuungsmaße

Mess-niveau	Eigenschaft	Kennzahl	Beschreibung
Nominal	gleich/ungleich	Lage: Modus	häufigster vorkommender Wert
Ordinal	größer/kleiner	Lage: Median und Quartile	teilen die Verteilung in Hälfte/Viertel
		Streuung: Quartilabstände	Abstände zwischen den Quartilen
Metrisch	gleiche Abstände (+/-)	Lage: Mittelwert	Durchschnitt
		Streuung: Standardabweichung	Durchschnittliche Abweichung um den Mittelwert
		Streuung: Spannweite	Breite der Verteilung (Abstand zw. kleinstem und größtem Wert)

Bedeutung des Messniveaus für die Kennzahlen

Modus:

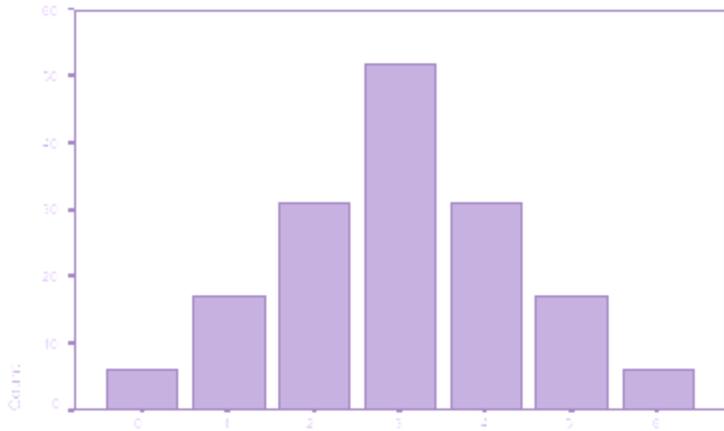
einziges Maß, das ab **nominalen Datenniveau** verwendet werden darf:
Vorraussetzung: keine, Interpretation oft nicht sinnvoll

Median, Quartile, Quartilabstände:

dürfen ab **ordinalen Datenniveau** verwendet werden:
Vorraussetzung: Messwerte haben eine angebbare **Reihenfolge**
von einer größten bis zu einer kleinsten Ausprägung

Mittelwert, Standardabweichung, Spannweite:

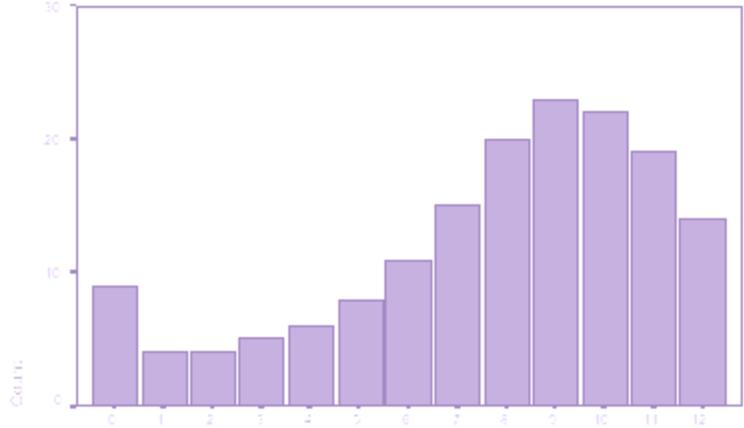
dürfen nur bei **metrischem Datenniveau** verwendet werden:
Vorraussetzung: Die Messwerte haben eine Maßeinheit (auch „Punkte“)
Die Abstände zwischen den Messwerten sind gleich



← „Anzahl Wörter Dolmetsch“

← „Arztbesuche im Jahr - Bevölkerung“

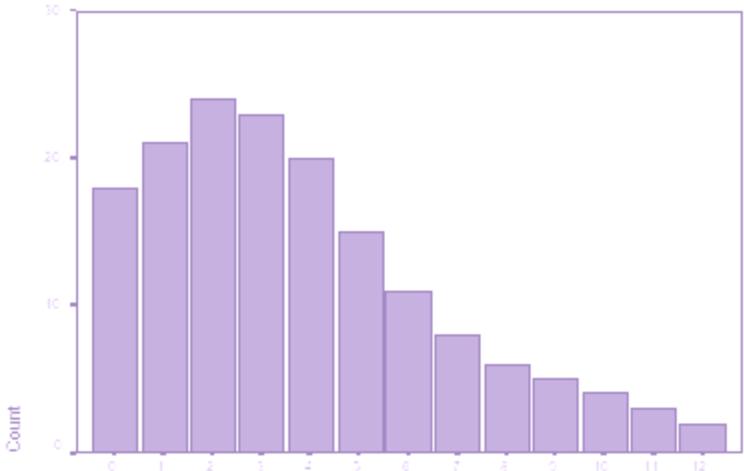
symmetrisch



← „Anzahl Wörter Berater:in“

← „Arztbesuche im Jahr – Chronisch Kranke“

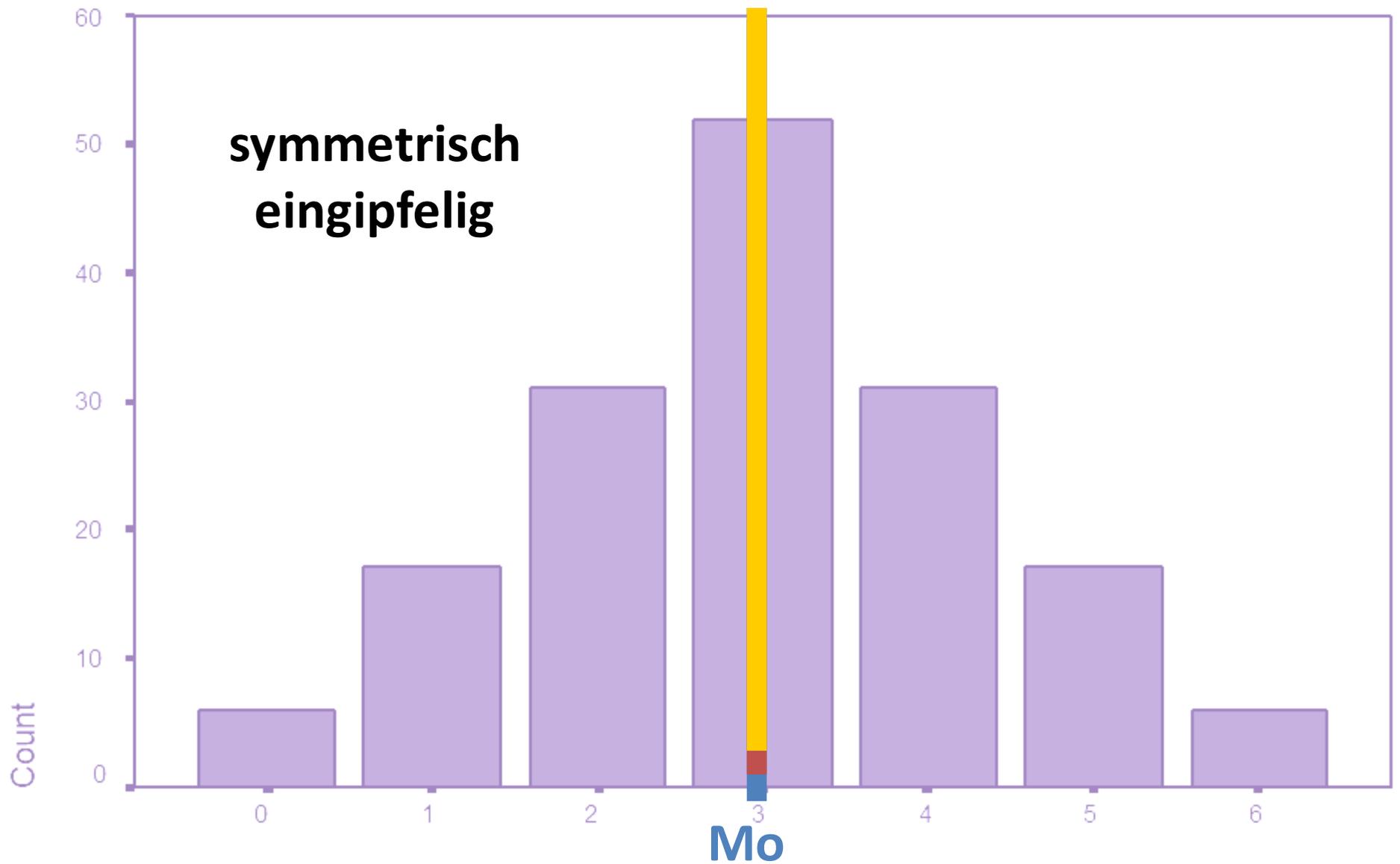
rechtsgipfelig



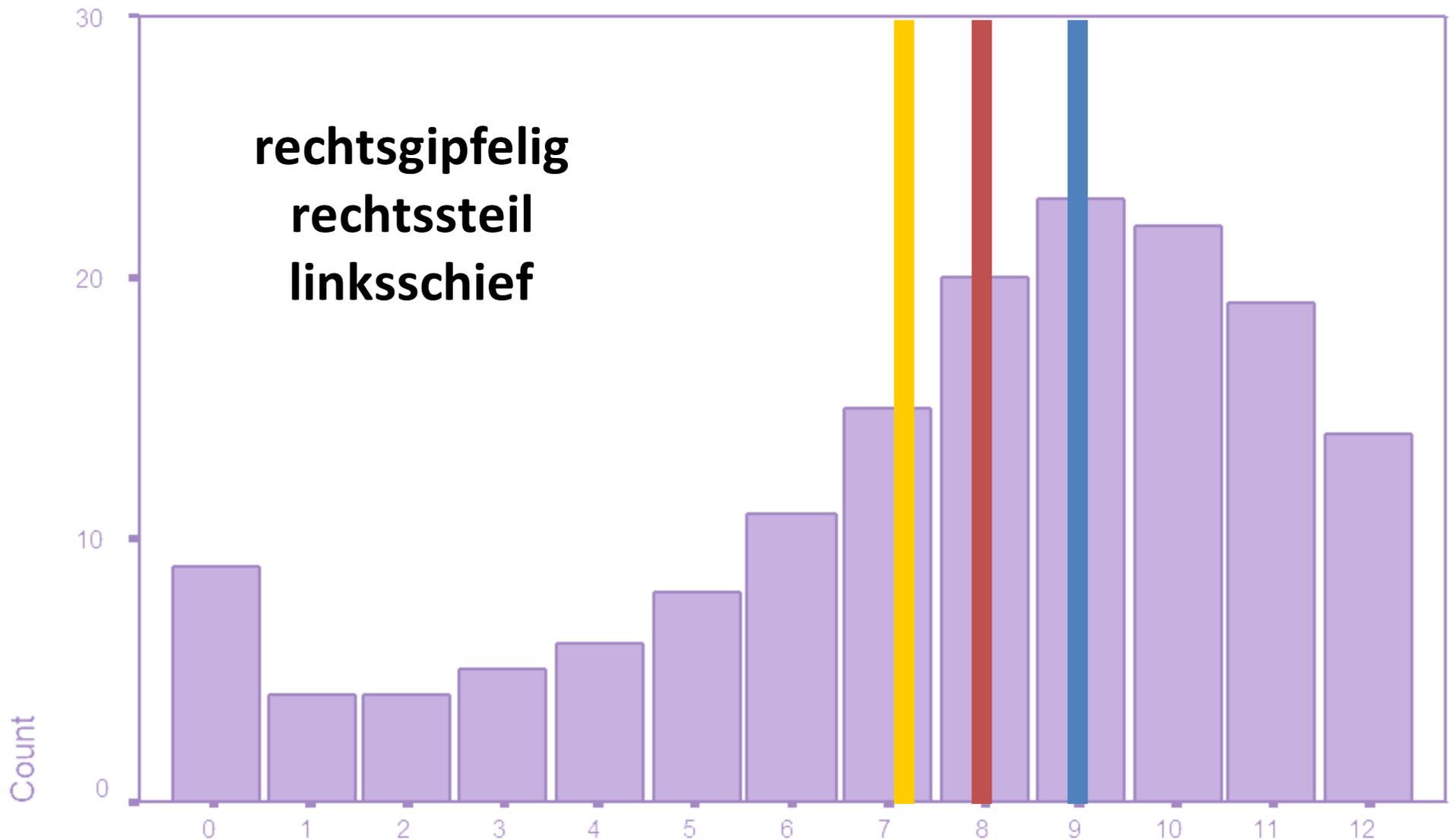
← „Anzahl Wörter Klient:in“

← „Arztbesuche im Jahr – Jugendliche“

linksgipfelig

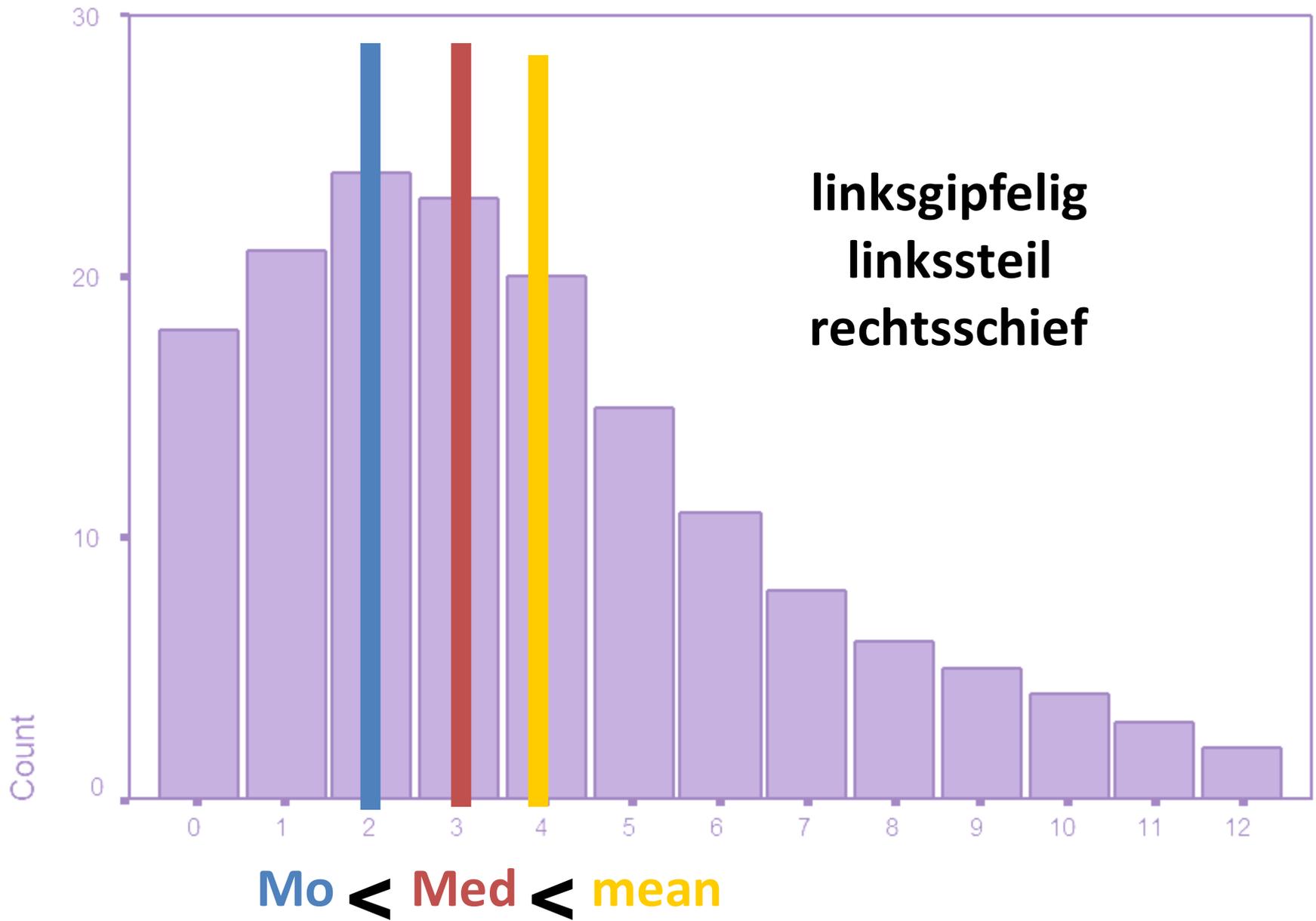


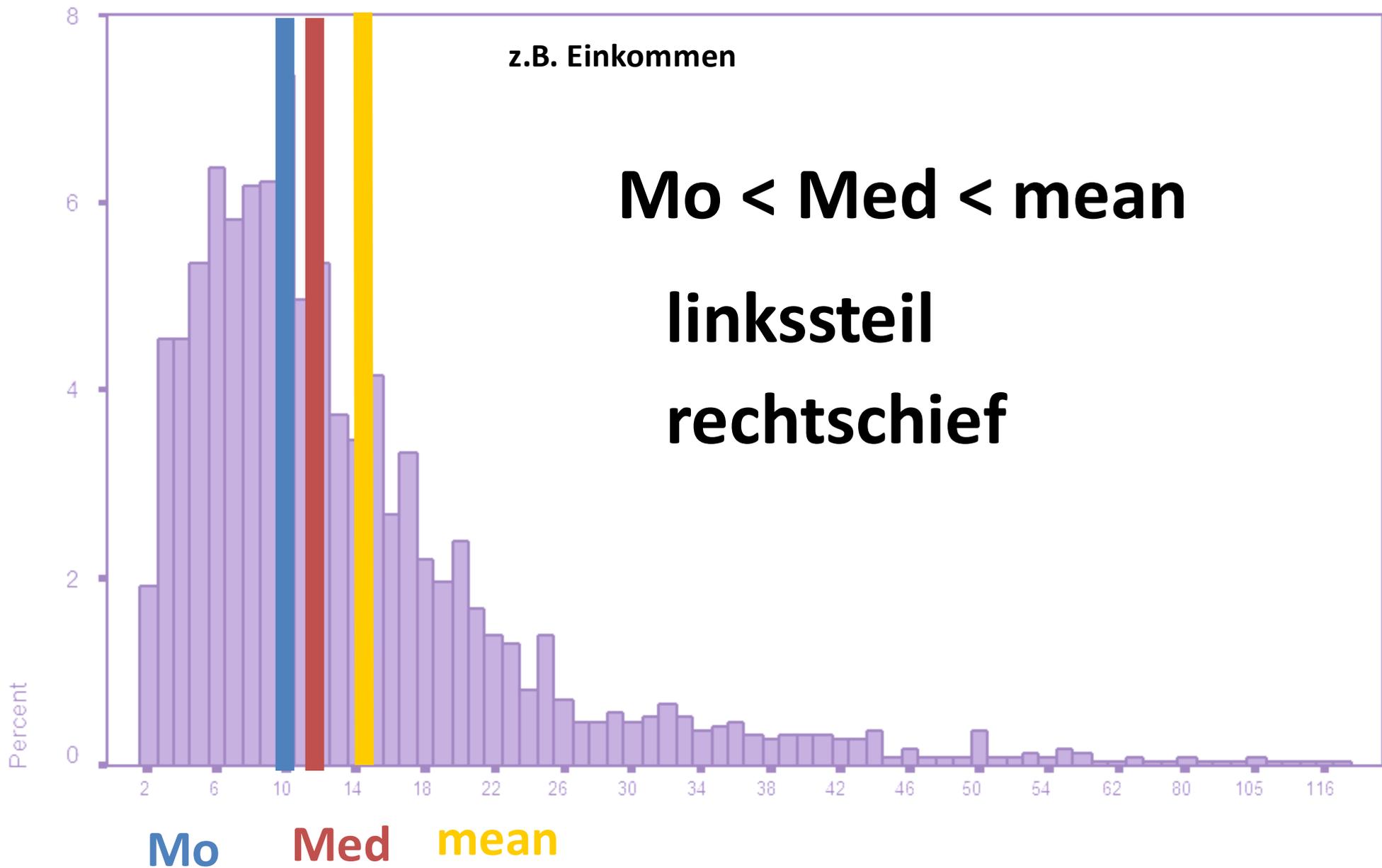
== Med
== mean

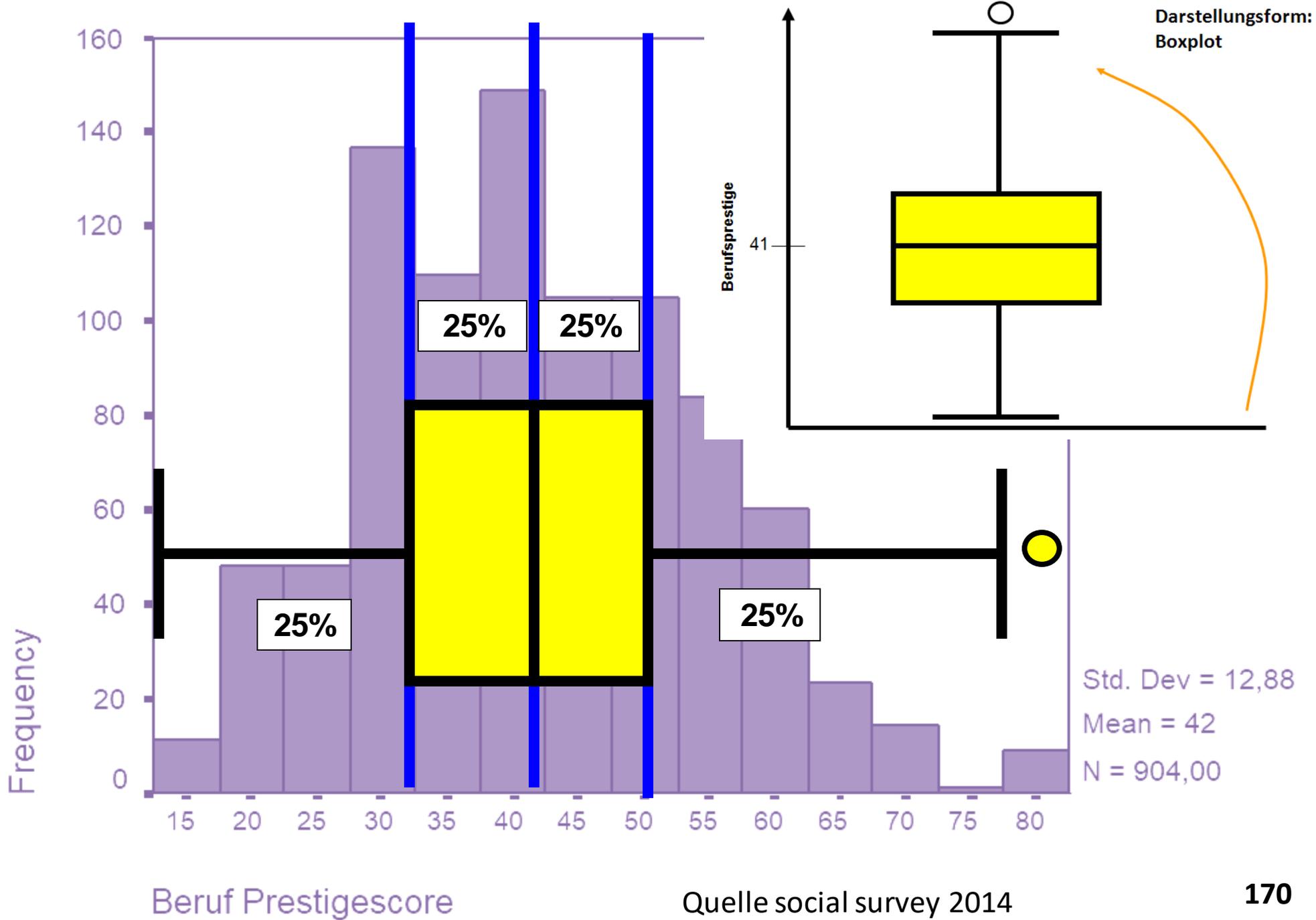


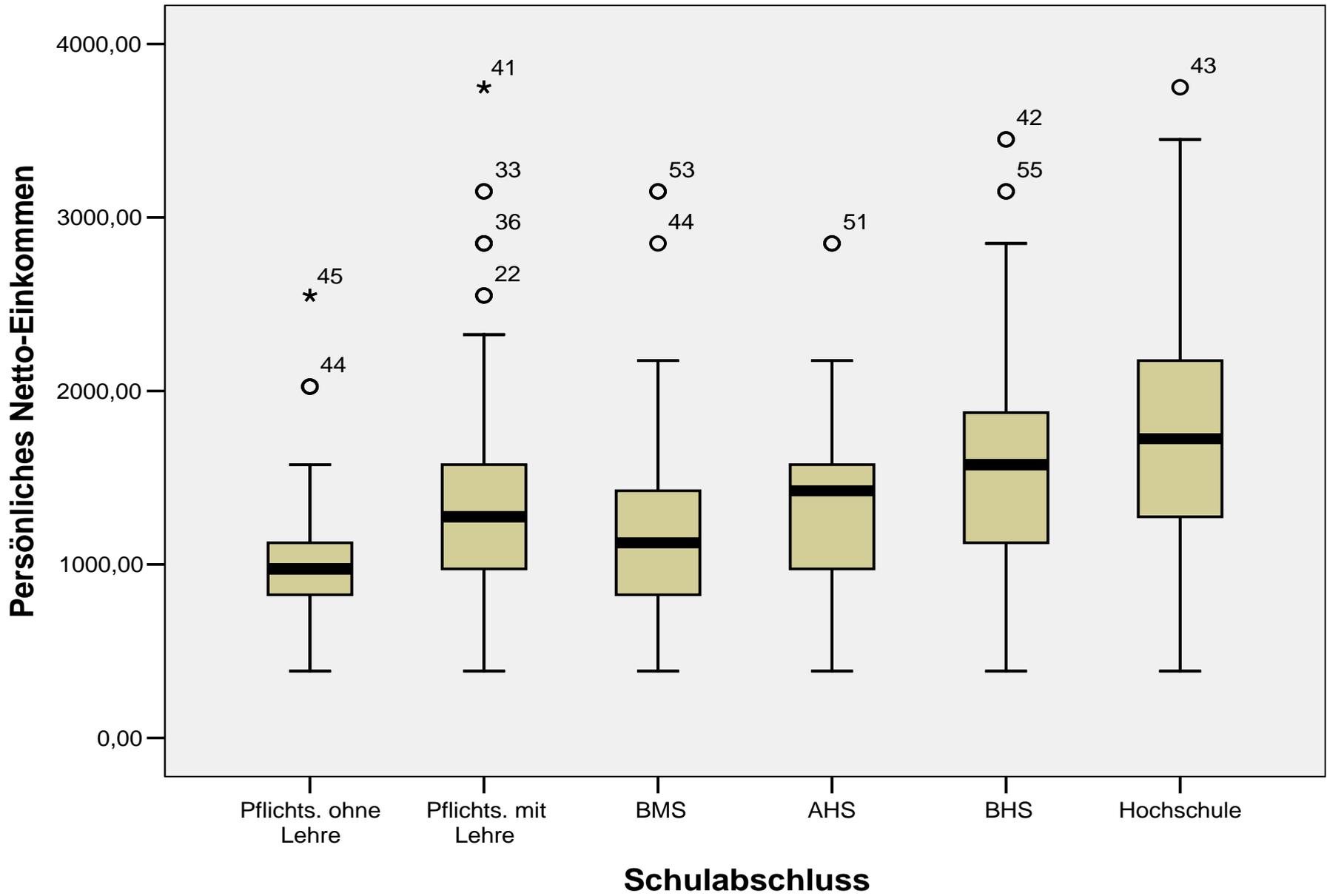
rechtsgipfelig
rechtssteil
linksschief

mean < Med < Mo



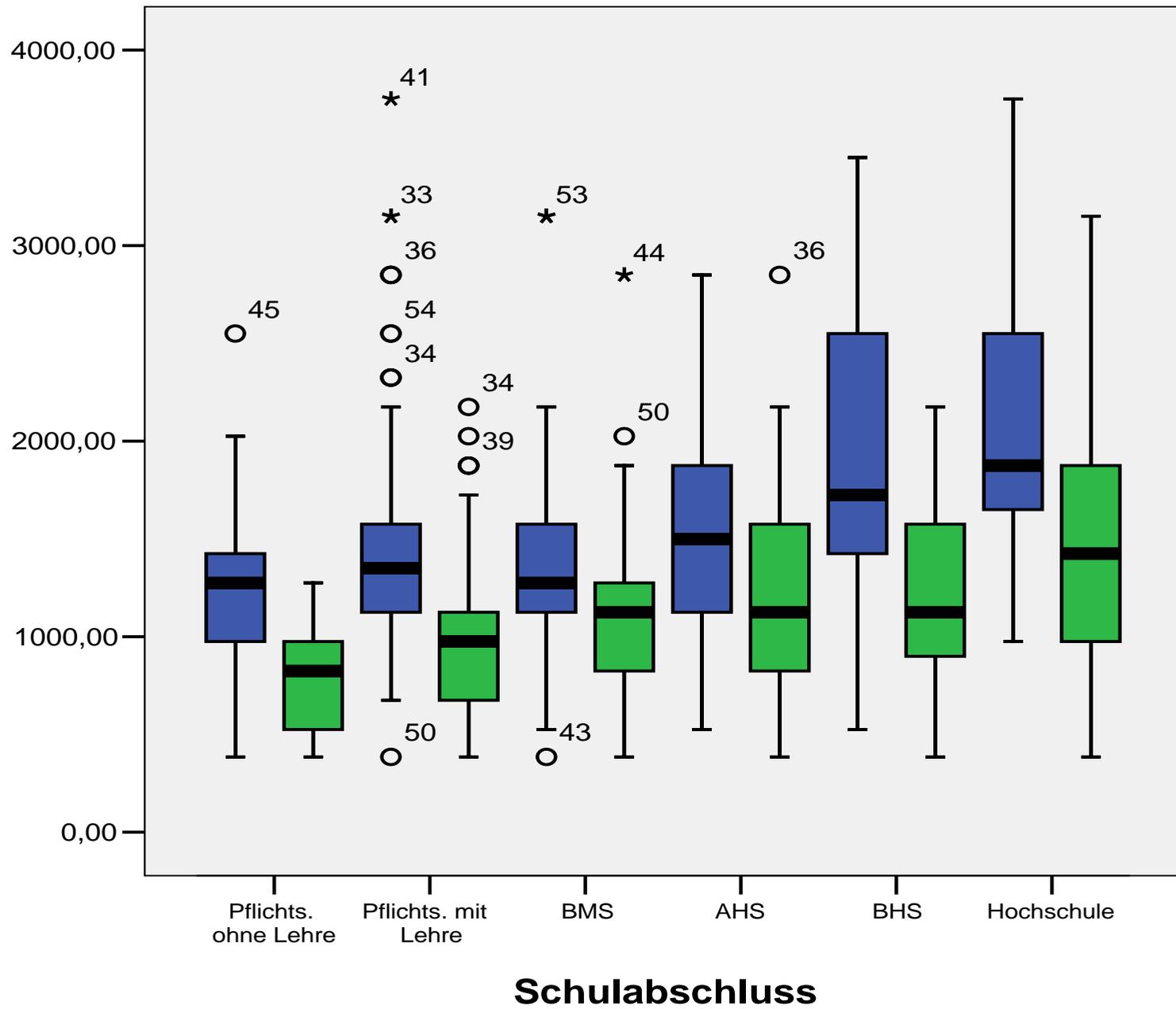






Cases weighted by Gewichtungsfaktor

Persönliches Netto-Einkommen



Geschlecht
■ männlich
■ weiblich

Cases weighted by Gewichtungsfaktor

Median

versus

Mittelwert

bei **ordinalen** und **metrischen** Variablen

nur bei **metrischem** Datenniveau

Anwendung bei **schiefer** Verteilung:
weil er immer innerhalb des Gipfels liegt
- von Ausreißern unbeeinflusst

Anwendung bei **symmetrischer** Verteilung:
- weil er sehr stark von Ausreißern aus dem Zentrum des Gipfels „gezogen“ wird

Je weiter Median und Mittelwert auseinander liegen, desto schiefer ist die Verteilung.

Liegen Median und Mittelwert sehr nahe, ist es eine symmetrische Verteilung.

Vorsicht vor manipulativer Datenbeschreibung!!

Dr. Göttlich beim Pharmakongress:
„Ich verschreibe der Hälfte meiner Patienten mehr als 8 Medikamente!“

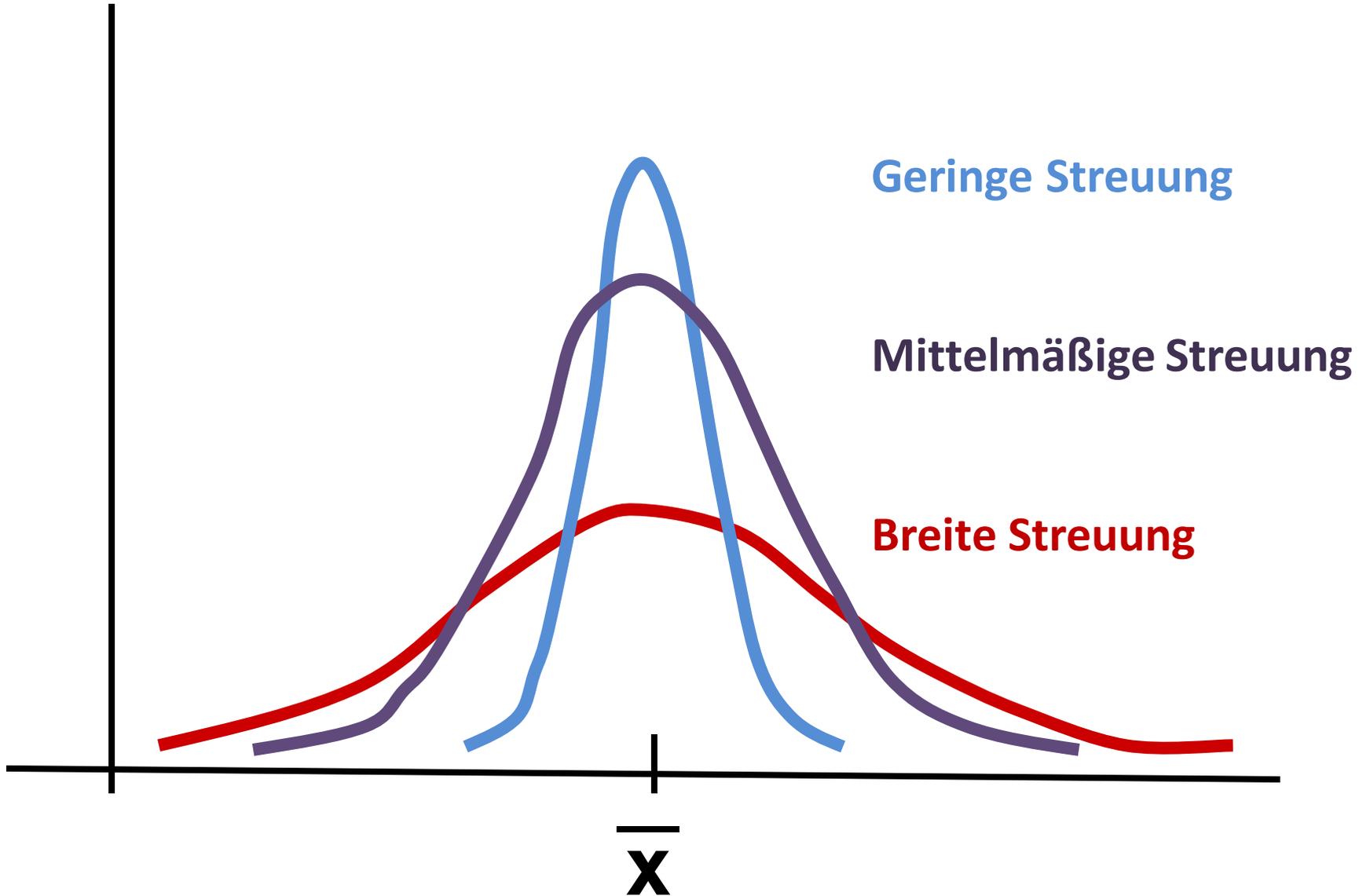
Dr. Göttlich beim Homöopathiekongress:
„Ich verschreibe durchschnittlich nur 7 Medikamente pro Patient.“

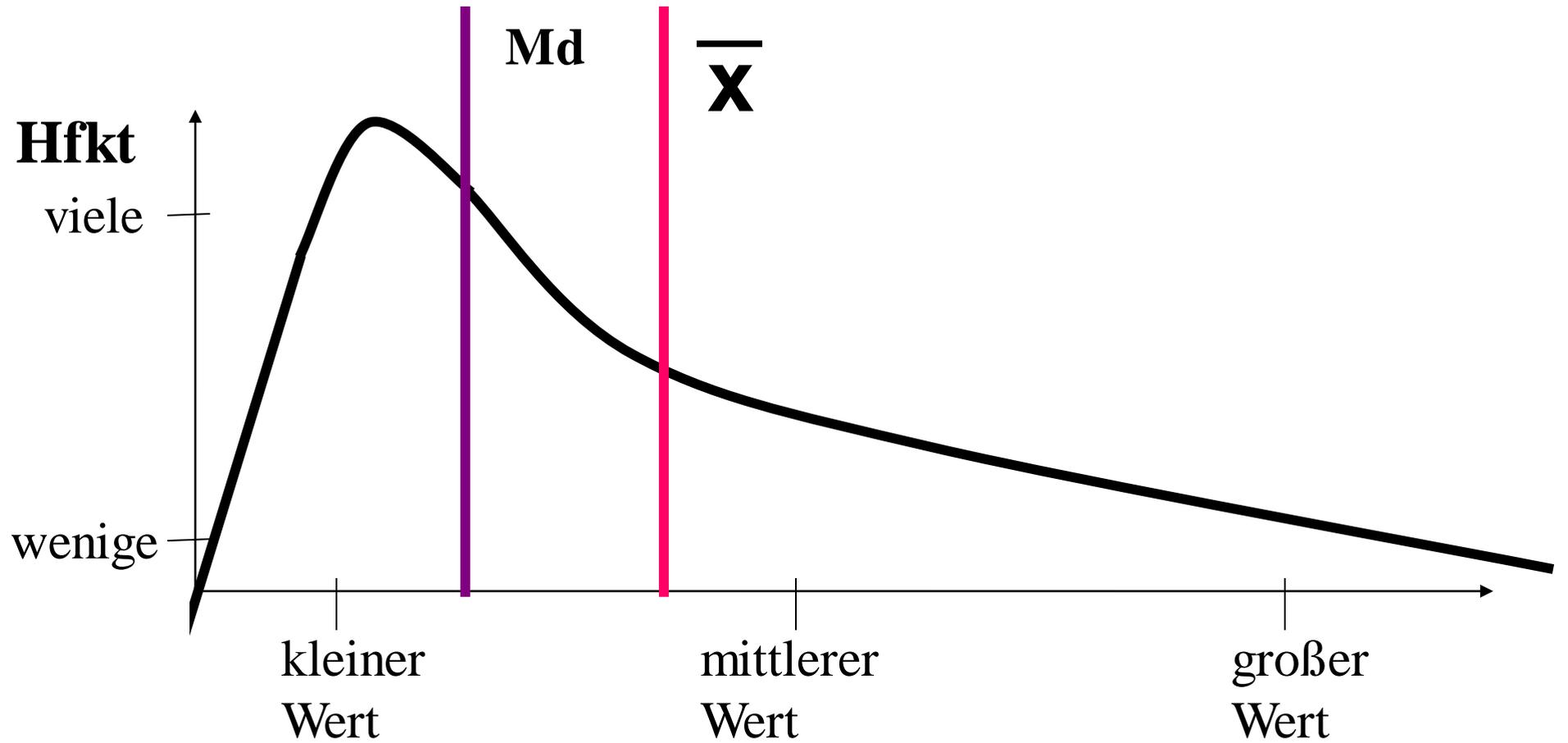
Median = „Realist“

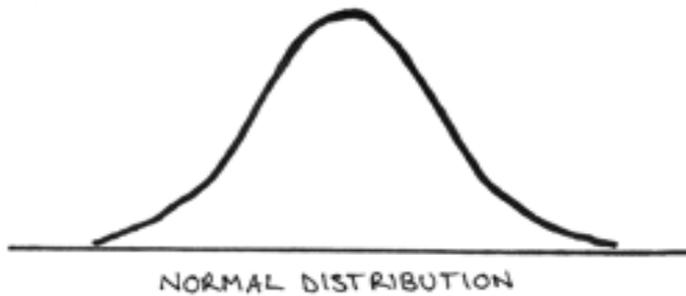
Mittelwert = „Nivellierer“

- verzerrt bei schiefer Verteilung das Ergebnis (z.B. Eink.)
- hat bei großer Streuung wenig Aussagekraft

Gleicher Mittelwert – unterschiedliche Streuung



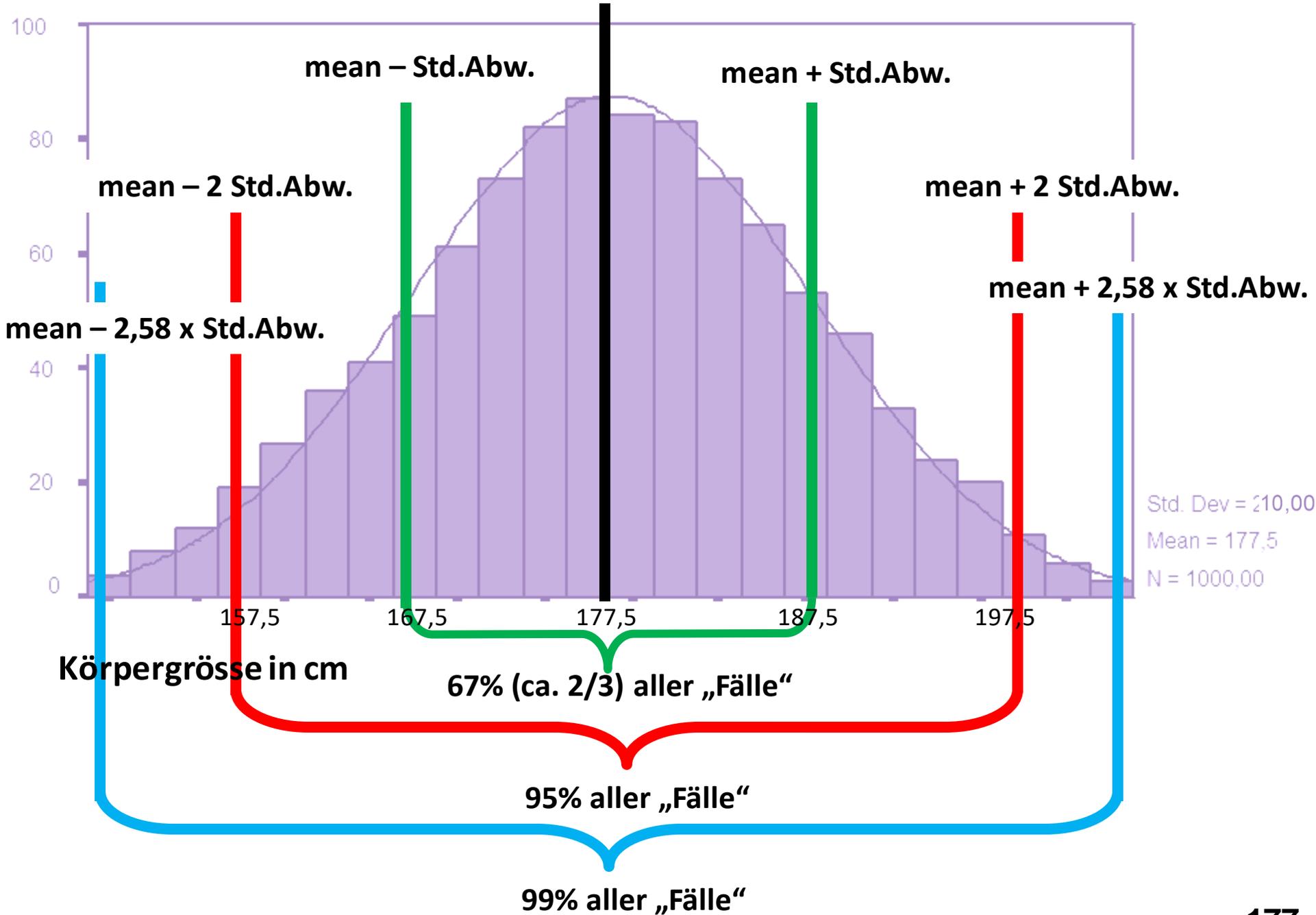


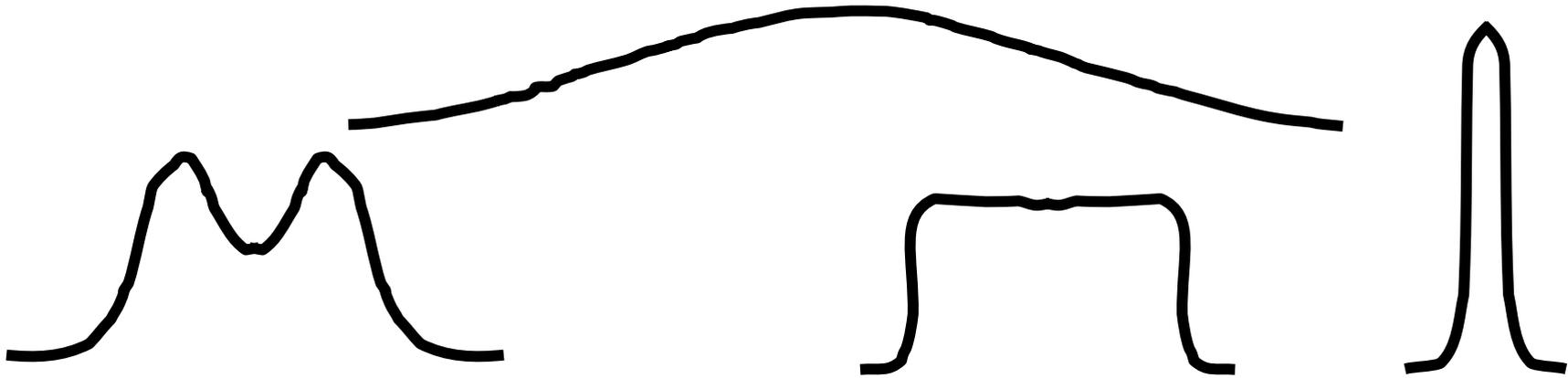


Normalverteilung Konfidenzintervall

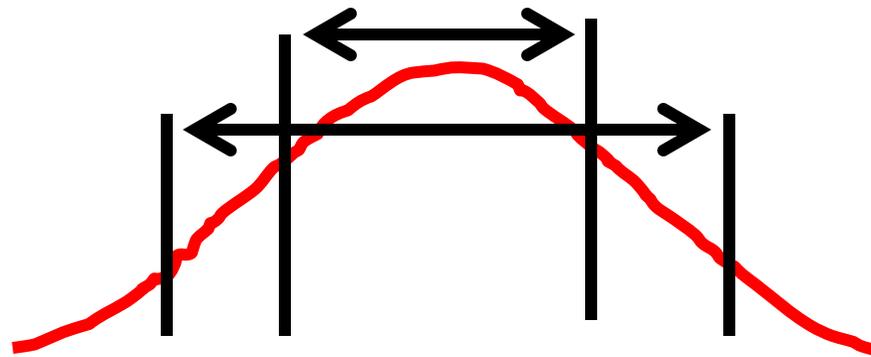


Fraun.





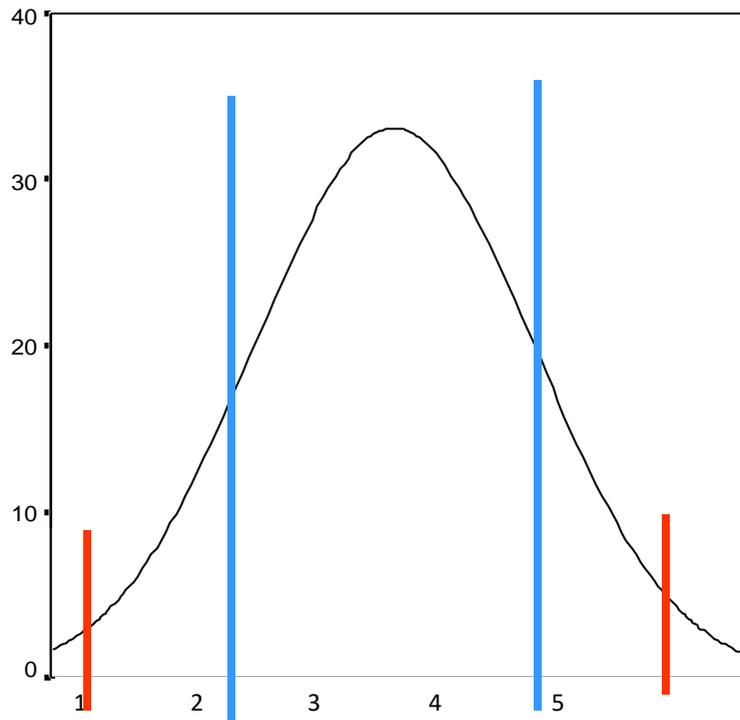
Symmetrische Verteilung: jede spiegelgleiche Form.



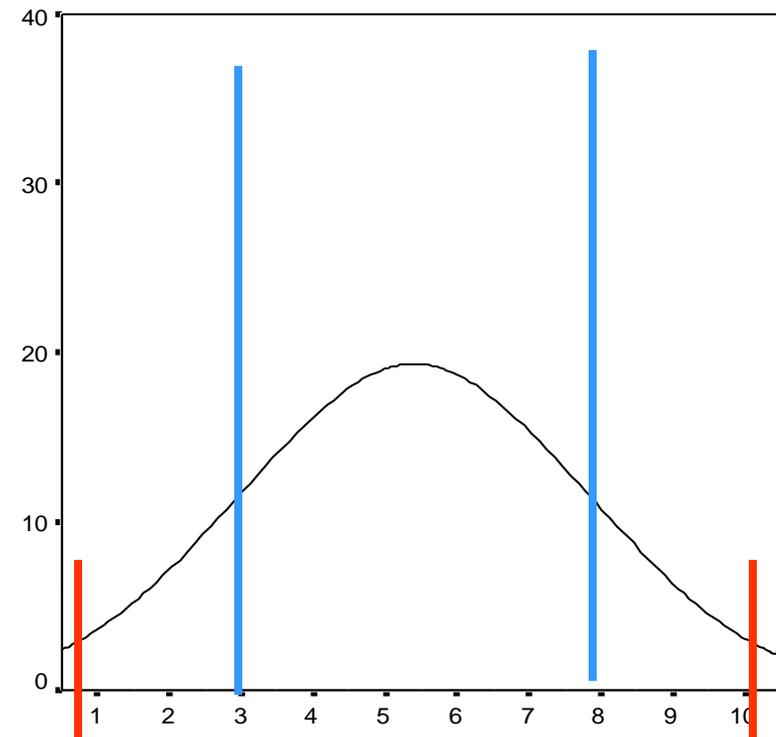
Normalverteilung - Glockenform: Intervallkriterien treffen zu.

Zufallsverteilung - Normalverteilung

Weniger n,
weniger Kategorien



Mehr n,
mehr Kategorien



Intervallkriterien treffen in jedem Fall zu.

Empirische Normalverteilung

empirisch = anhand der vorliegenden Stichprobe erhoben

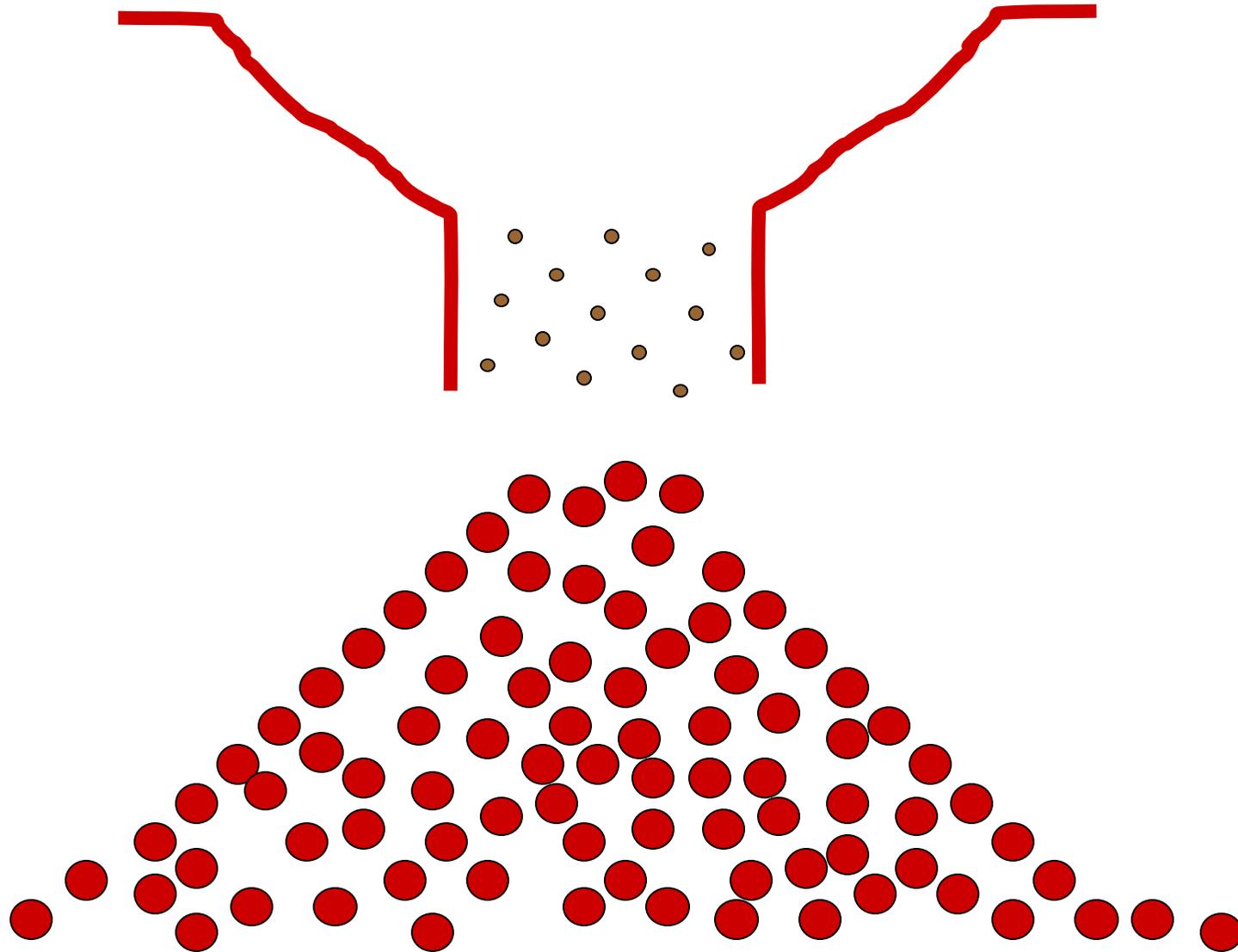
Normalverteilung = Verteilung von zufällig zustande gekommenen Merkmalen

Das heißt: Merkmale, deren Entstehungsfaktoren voneinander unabhängig sind (keine von Menschen geschaffene Gesetzmäßigkeit) Das sind:

1. „**Natürliche**“ **Merkmale** (Körpergröße, Körpergewicht, ...)
2. Neu berechnete Variablen (**Indizes**),
die sich aus mehreren Items zusammensetzen (z.B. IQ)

Eigenschaften der Normalverteilung

1. **symmetrisch**: Median = Mittelwert
2. **Glockenform**: Intervallkriterien treffen immer zu:
Im Bereich zwischen $MW \pm 1$ x Standardabweichung liegen 66,67% der Verteilung.
Im Bereich zwischen $MW \pm 2$ x Std.Abw. (genau: **1,96** x) liegen 95% der Verteilung.
Im Bereich zwischen $MW \pm 2,58$ x Std.Abw. liegen 99% der Verteilung.



Logik der schließenden Statistik

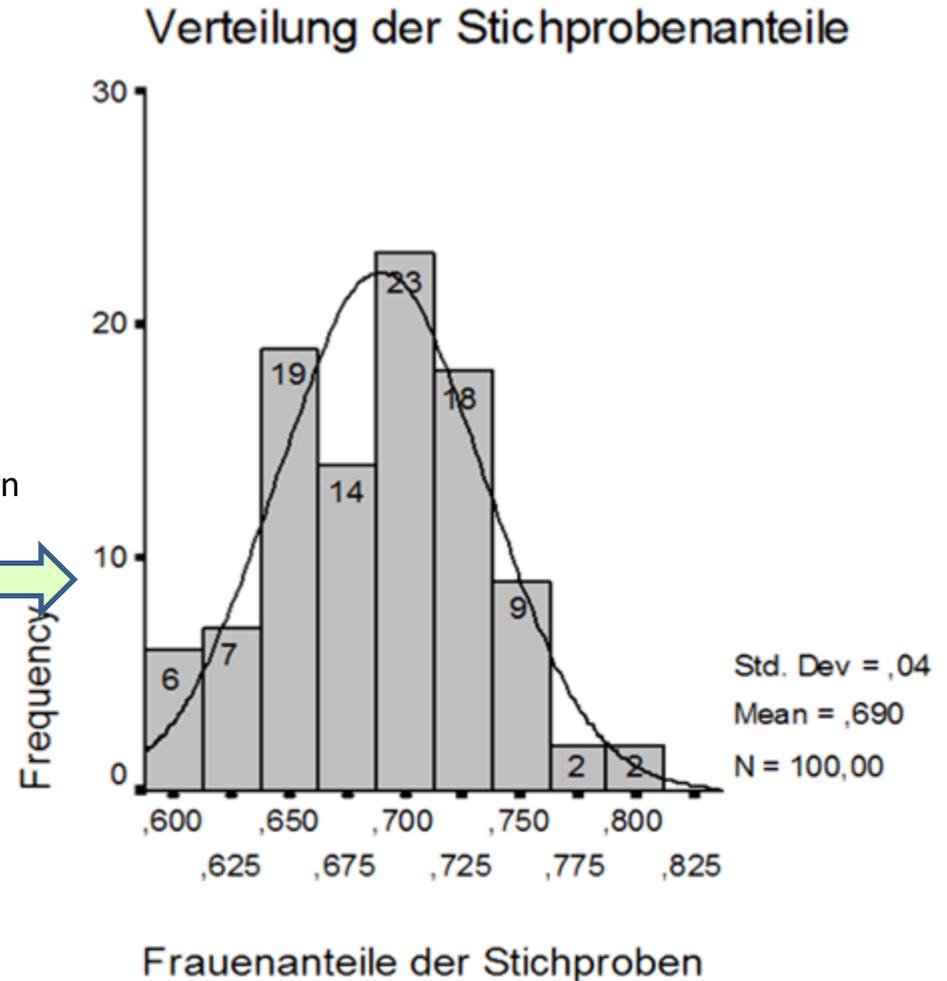
- Man stellt sich vor, unendlich viele Stichproben zu ziehen.
- Misst man ein Merkmal unendlich oft, ähnelt die Verteilung am Ende einer Normalverteilung
- Die Standardabweichung dieser Verteilung nennt man Standardfehler.

Beispiel

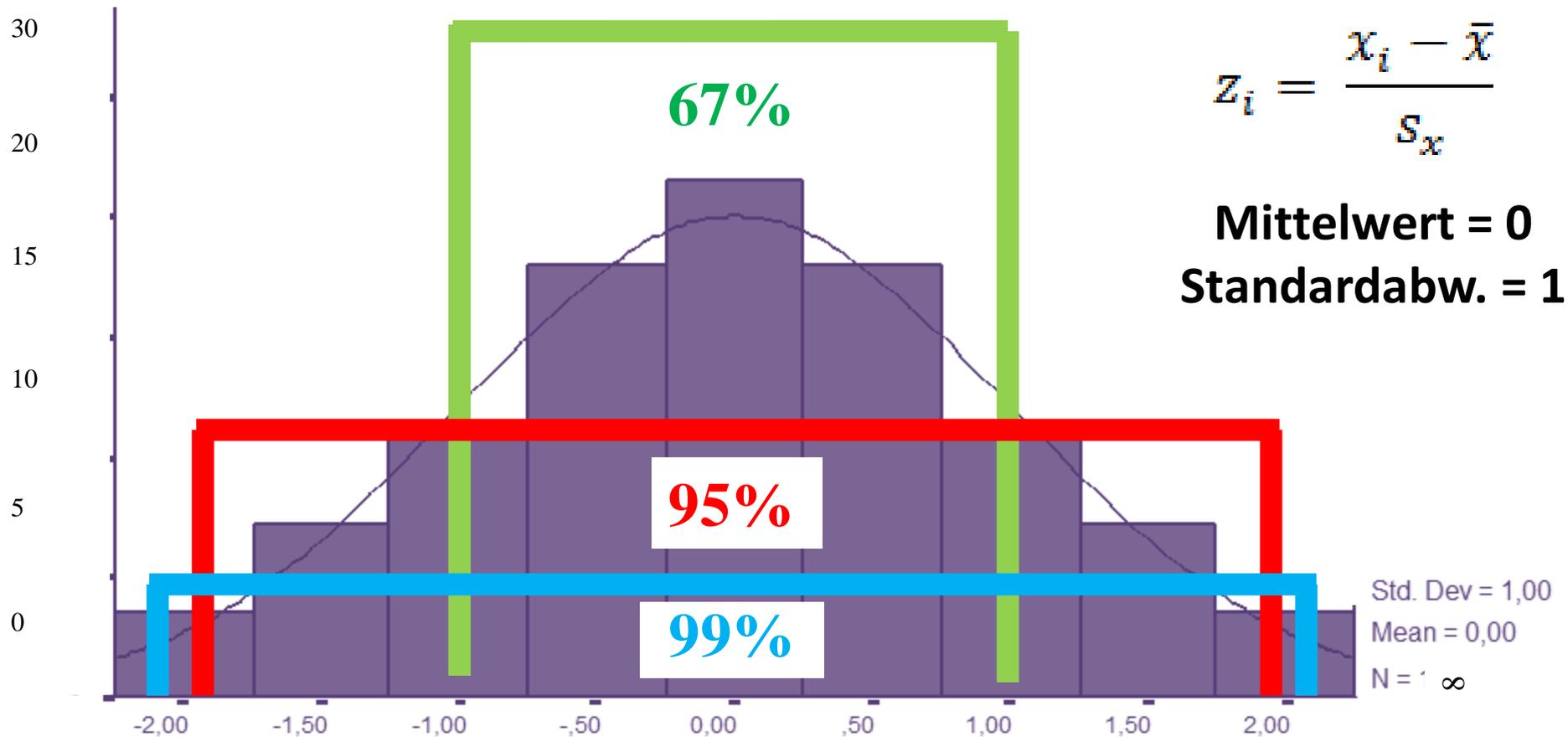
Wie hoch ist der Frauenanteil unter den Studierenden?

1. Stichprobe: 65%
2. Stichprobe: 73%
3. Stichprobe: 71%
- ...
30. Stichprobe: 69%
31. Stichprobe: 61%
- ...
500. Stichprobe: 64%
501. Stichprobe: 72%
502. Stichprobe: 71%
-
- Usw.

nach 100 gezogenen Stichproben



Bei unendlich oft gezogenen Stichproben =
Annäherung an eine Normalverteilung!



Mittelwert \pm Std.Abw. \rightarrow 67% (ca. 2/3) aller Fälle

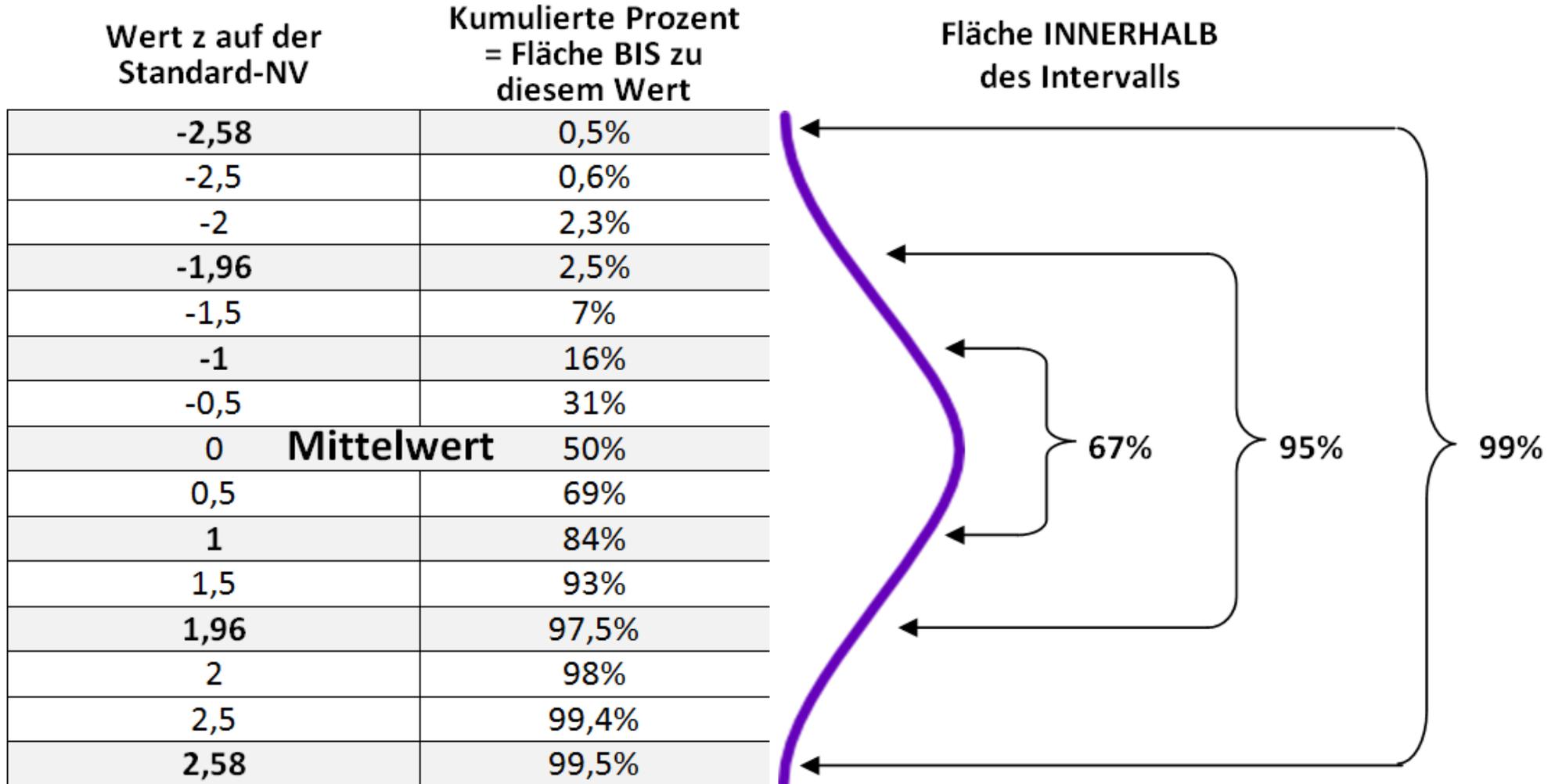
Mittelwert $\pm 1,96$ *Std.Abw. \rightarrow 95% (ca. 2/3) aller Fälle

Mittelwert $\pm 2,58$ *Std.Abw. \rightarrow 99% (ca. 2/3) aller Fälle

Standardnormalverteilung

Mittelwert = 0
Standardabw. = 1

185



Berechnen des Konfidenzintervalls

$$\text{KI}_{95\%} = \left[\text{Anteilswert} \pm 1,96 \times \text{Standardfehler des Anteilswertes} \right]$$

p

$$S_p = \sqrt{\frac{p^*(1-p)}{n}}$$

Berechnen des Konfidenzintervalls – für den Mittelwert

$$\text{KI}_{95\%} = \left[\text{Mittelwert} \pm 1,96 \times \text{Standardfehler des Mittelwertes} \right]$$

\bar{x}

$$S_{\bar{x}} = \sqrt{\frac{S_x^2}{n}} = \frac{S_x}{\sqrt{n}}$$

Berechnen des Konfidenzintervalls für einen **Mittelwert**

(= Schätzintervall des „wahren“ **Mittelwertes** in der GG)

$$KI_{95\%} \text{ Obergrenze} = \bar{x} + 1,96 * s_{\bar{x}}$$
$$KI_{95\%} \text{ Untergrenze} = \bar{x} - 1,96 * s_{\bar{x}}$$

$$KI_{95\%} = \text{Mittelwert} \pm 1,96 * \text{Standardfehler}$$

$$KI_{99\%} \text{ Obergrenze} = \bar{x} + 2,58 * s_{\bar{x}}$$
$$KI_{99\%} \text{ Untergrenze} = \bar{x} - 2,58 * s_{\bar{x}}$$

$$KI_{99\%} = \text{Mittelwert} \pm 2,58 * \text{Standardfehler}$$

Berechnen des Konfidenzintervalls für einen Anteilswert

(= Schätzintervall des „wahren“ Anteils in der GG)

$$KI_{95\%} \text{ Obergrenze} = p + 1,96 * S_p$$

$$KI_{95\%} \text{ Untergrenze} = p - 1,96 * S_p$$

$$KI_{95\%} = \text{Anteilswert} \pm 1,96 * \text{Standardfehler}$$

$$KI_{99\%} \text{ Obergrenze} = p + 2,58 * S_p$$

$$KI_{99\%} \text{ Untergrenze} = p - 2,58 * S_p$$

$$KI_{99\%} = \text{Anteilswert} \pm 2,58 * \text{Standardfehler}$$

Theoretische Normalverteilung = Wahrscheinlichkeitsverteilung

theoretisch = mithilfe von Wahrscheinlichkeitsrechnung berechenbar

Verwendung:

→ Berechnung von **Konfidenzintervallen** (=Schwankungsbreiten) von Stichprobenergebnissen: mithilfe der Standardnormalverteilung wird der wahre Parameter in der GG geschätzt.

→ WS-Bestimmung eines Prüfmaß zur **Hypothesenentscheidung**: Bestimmen des α -Fehlers. Die Normalverteilung entspricht der zufälligen Verteilung eines Prüfmaßes bei gegebener Unabhängigkeit (=Zufall).

Eigenschaften der Standard-Normalverteilung: wie jede Normalverteilung, und:

1. **Mittelwert = 0**

2. **Standardabweichung = 1:** Intervallkriterien treffen immer zu:

Im Bereich zwischen $MW \pm 1$ x Standardabweichung liegen 66,67% der Verteilung.

Im Bereich zwischen $MW \pm 2$ x Std.Abw. (genau: **1,96** x) liegen 95% der Verteilung.

Im Bereich zwischen $MW \pm$ **2,58** x Std.Abw. liegen 99% der Verteilung.

Kreuztabellen

Balkendiagramme

Häufigkeitstabelle Wortanzahl ÄrztIn in 200er-Schritten

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 400-600	3	4,0	8,1	8,1
	2 601-800	3	4,0	8,1	16,2
	3 801-1000	1	1,3	2,7	18,9
	4 1001-1200	1	1,3	2,7	21,6
	5 1201-1400	2	2,7	5,4	27,0
	6 1401-1600	3	4,0	8,1	35,1
	7 1601-1800	4	5,3	10,8	45,9
	8 1801-2000	2	2,7	5,4	51,4
	9 2001-2200	4	5,3	10,8	62,2
	10 2201-2400	2	2,7	5,4	67,6
	11 2401-2600	2	2,7	5,4	73,0
	12 2601-2800	1	1,3	2,7	75,7
	13 2801-3000	2	2,7	5,4	81,1
	14 3001-3200	1	1,3	2,7	83,8
	15 3201-3400	1	1,3	2,7	86,5
	16 3401-3600	1	1,3	2,7	89,2
	17 3601-3800	2	2,7	5,4	94,6
	18 3801-4000	1	1,3	2,7	97,3
	19 4001-4200	1	1,3	2,7	100,0
	Gesamt	37	49,3	100,0	
Fehlend	99 keine Angabe	38	50,7		
Gesamt		75	100,0		

Aufteilen in
drei Kategorien

„dichotomisieren“
= Aufteilen in zwei
Kategorien

Syntax der Umkodierung

```

RECODE
  wort_A_in200 (99=99) (1 thru 6=1) (7 thru 10=2) (11 thru 19=3) INTO wort_A_kat .
EXECUTE .
VARIABLE LABELS wort_A_kat ,Wortanzahl Ärztin'.
EXECUTE .
  
```

wort_A_kat Wortanzahl Ärztin

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 400-1600	13	17,3	35,1	35,1
	2 1601-2400	12	16,0	32,4	67,6
	3 2401-4200	12	16,0	32,4	100,0
	Gesamt	37	49,3	100,0	
Fehlend	99 keine Angabe	38	50,7		
Gesamt		75	100,0		

Kreuztabelle = Zweidimensionale Häufigkeitstabelle

	Merkmal 2		
Merkmal 1	y_1	y_2	Summe
x_1	n_{11}	n_{12}	$n_{1.}$
x_2	n_{21}	n_{22}	$n_{2.}$
x_3	n_{31}	n_{32}	$n_{3.}$
Summe	$n_{.1}$	$n_{.2}$	n

Eine der gebräuchlichsten Analysemethoden –
bei Nominalskalen und allen kategorisierten Skalen (Übersichtlichkeit!)
Achtung: Je nach Kategorisierung unterschiedliche Ergebnisse

Kreuztabelle

„Die Erdäpfel/Kartoffeln müssen noch geschält werden.“

↓ **Merkmal B: Spalten**

		gruppe Gruppe		
		1 SchülerInnen	2 LehrerInnen	Gesamt
v13_23 Ausdrücke:	1 Erdäpfel	78	14	92
Erdäpfel / Kartoffel	2 Kartoffel	162	9	171
Gesamt		240	23	263

Merkmal A: Zeilen

Kreuztabelle

↓ **Merkmal B: Spalten**

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

Anzahl



		sex_A Geschlecht ÄrztIn		
		0 männlich	1 weiblich	Gesamt
wort_A Wortanzahl ÄrztIn	1 bis zu 1900 Wörter	16	2	18
	2 mehr als 1901 Wörter	10	9	19
Gesamt		26	11	37

Merkmal A: Zeilen

Unterscheiden sich die Gespräche hinsichtlich der Wortanzahl der ÄrztInnen und dem Geschlecht des/der ÄrztIn?

Verarbeitete Fälle

	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
wort_A_kat Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn	37	49,3%	38	50,7%	75	100,0%

wort_A_kat Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

				sex_A Geschlecht ÄrztIn		Gesamt
				0 männlich	1 weiblich	
wort_A_kat Wortanzahl ÄrztIn	1 400-1600	Anzahl	12	1	13	
		% von sex_A Geschlecht ÄrztIn	46,2%	9,1%	35,1%	
	2 1601-2400	Anzahl	7	5	12	
	% von sex_A Geschlecht ÄrztIn	26,9%	45,5%	32,4%		
	3 2401-4200	Anzahl	7	5	12	
	% von sex_A Geschlecht ÄrztIn	26,9%	45,5%	32,4%		
Gesamt		Anzahl	26	11	37	
		% von sex_A Geschlecht ÄrztIn	100,0%	100,0%	100,0%	

Unterscheiden sich die Gespräche hinsichtlich der Wortanzahl der ÄrztInnen und dem Geschlecht des/der ÄrztIn?

Analysieren → Deskriptive Statistiken → Kreuztabellen... wort_A + sex_A
 Zellen → beobachtete Werte und **Spaltenprozentwerte** → weiter → ok

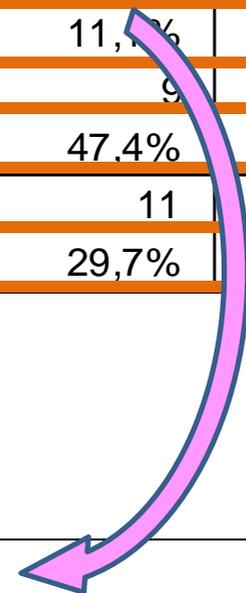
wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A	bis zu 1900 Wörter	Anzahl	16	2	18
Wortanzahl		%	61,5%	18,2%	48,6%
ÄrztIn	mehr als 1901 Wörter	Anzahl	10	9	19
		%	38,5%	81,8%	51,4%
Gesamt		Anzahl	26	11	37
		%	100,0%	100,0%	100,0%

Analysieren → Deskriptive Statistiken → Kreuztabellen... wort_A + sex_A
 Zellen → beobachtete Werte und **Zeilenprozentwerte** → weiter → ok

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A	bis zu 1900 Wörter	Anzahl	16	2	18
Wortanzahl		%	88,9%	11,1%	100,0%
ÄrztIn	mehr als 1901 Wörter	Anzahl	10	9	19
		%	52,6%	47,4%	100,0%
Gesamt		Anzahl	26	11	37
		%	70,3%	29,7%	100,0%



Schwankungsbreite (Zufallsfehler): ca. 15% → 0 % bis 25 %

Daten gewichten.....

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A	bis zu 1900 Wörter	Anzahl	160	20	180
Wortanzahl		%	88,9%	11,1%	100,0%
ÄrztIn	mehr als 1901 Wörter	Anzahl	100	90	190
		%	52,6%	47,4%	100,0%
Gesamt		Anzahl	260	110	370
		%	70,3%	29,7%	100,0%

Schwankungsbreite (Zufallsfehler): ca. 5% → 5% bis 15%

Analysieren → Deskriptive Statistiken → Kreuztabellen... wort_A + sex_A
 Zellen → beobachtete Werte und **Gesamtprozentwerte** → weiter → ok

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A	bis zu 1900 Wörter	Anzahl	16	2	18
Wortanzahl		% der Gesamtzahl	43,2%	5,4%	48,6%
ÄrztIn	mehr als 1901 Wörter	Anzahl	10	9	19
		% der Gesamtzahl	27,0%	24,3%	51,4%
Gesamt		Anzahl	26	11	37
		% der Gesamtzahl	70,3%	29,7%	100,0%

Berechnen der Prozentwerte bei einer Kreuztabelle

Bsp. Gespräche bis zu 1900 ARZT-Wörter bei Ärzten

Spalten% 16 Gespräche bis zu 1900 ARZT-Wörter von allen Ärzten:

$$\frac{n_1}{n_{col}} 100 = \frac{16}{26} 100 = 61,5\%$$

In 61,5% der Gespräche von Ärzten wurden von diesen bis zu 1900 Wörter gesprochen.

Zeilen % 16 Gespräche von Ärzten von allen Gesprächen bis zu 1900 Wörtern:

$$\frac{n_1}{n_{row}} 100 = \frac{16}{18} 100 = 88,9\%$$

90% der Gespräche bis zu 1900 ARZT-Wörtern wurden von Ärzten geführt.

Gesamt % 16 Gespräche von Ärzten mit bis zu 1900 Wörtern von allen Gesprächen insgesamt:

$$\frac{n_1}{g} 100 = \frac{16}{37} 100 = 43,2\%$$

Insgesamt sind 43 % aller Gespräche solche, wo männliche Ärzte bis zu 1900 Wörter gesprochen haben.

Analysieren → Deskriptive Statistiken → Kreuztabellen... dauer_kat + sex_A
 Zellen → nur Spaltenprozentwerte → weiter → ok
 → mit Strg+c ins Excel kopieren

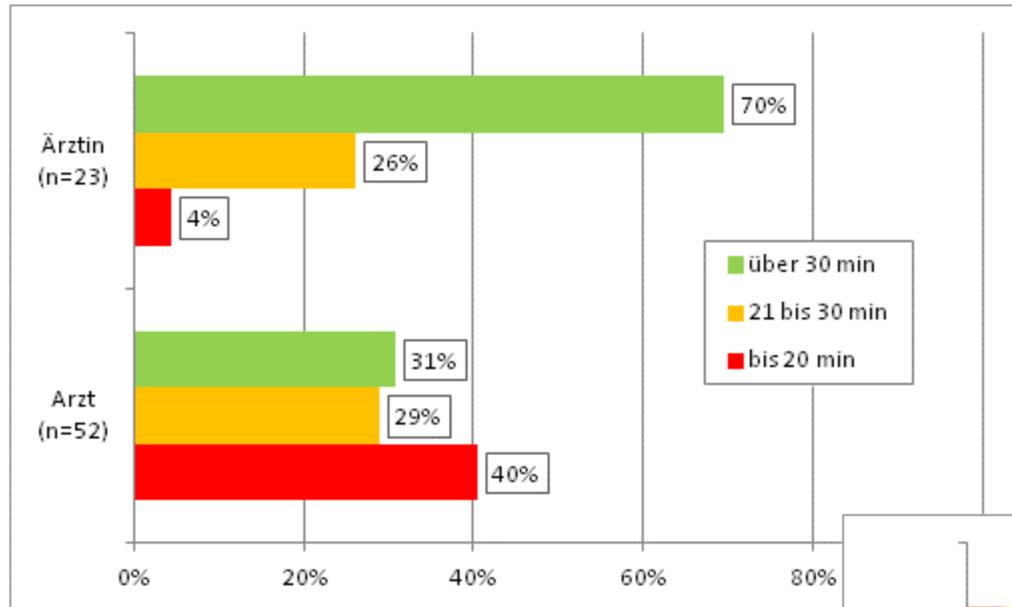
dauer_kat Gesprächsdauer * sex_A Geschlecht ÄrztIn Kreuztabelle

% von dauer_kat Gesprächsdauer

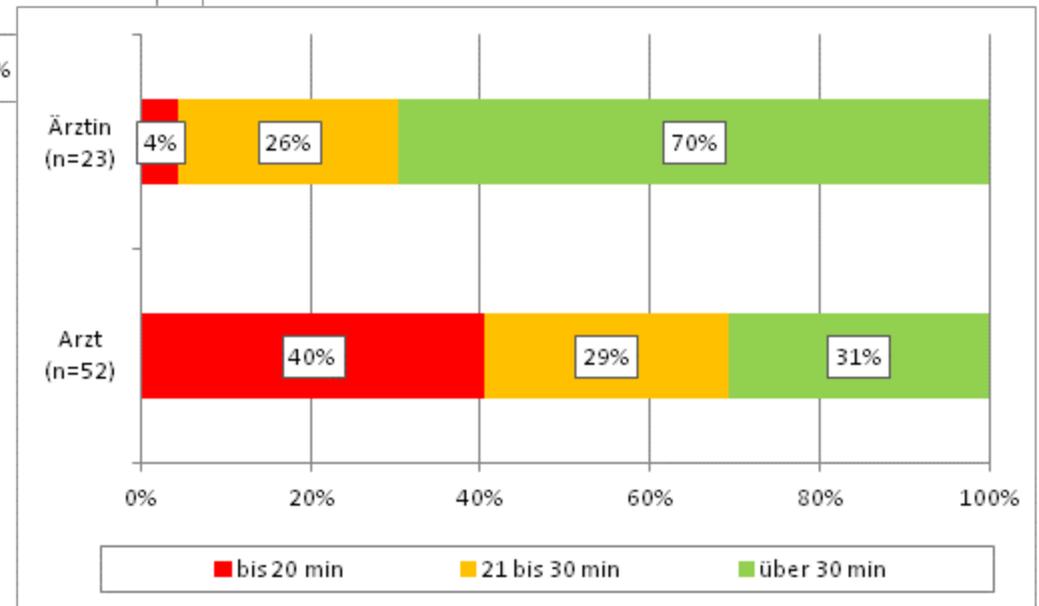
		sex_A Geschlecht ÄrztIn		Gesamt
		0 männlich	1 weiblich	
dauer_kat	1 bis 20 min	95,5%	4,5%	100,0%
Gesprächsdauer	2 21 bis 30 min	71,4%	28,6%	100,0%
	3 über 30 min	50,0%	50,0%	100,0%
Gesamt		69,3%	30,7%	100,0%

	A	B	C	D	E	F
1						
2	dauer_kat Gesprächsdauer * sex_A Geschlecht ÄrztIn Kreuztabelle					
3	% von dauer_kat Gesprächsdauer					
4			sex_A Geschlecht Ärz		Gesamt	
5			Arzt	Ärztin		
6	dauer_kat G	1 bis 20 min	95,5	4,5	100	
7		2 21 bis 30 m	71,4	28,6	100	
8		3 über 30 mi	50,0	50,0	100	
9		Gesamt	69,3	30,7	100	
10						

Gesprächsdauer nach Geschlecht der/des Ärzt:in → Ärzt:in ist 100%



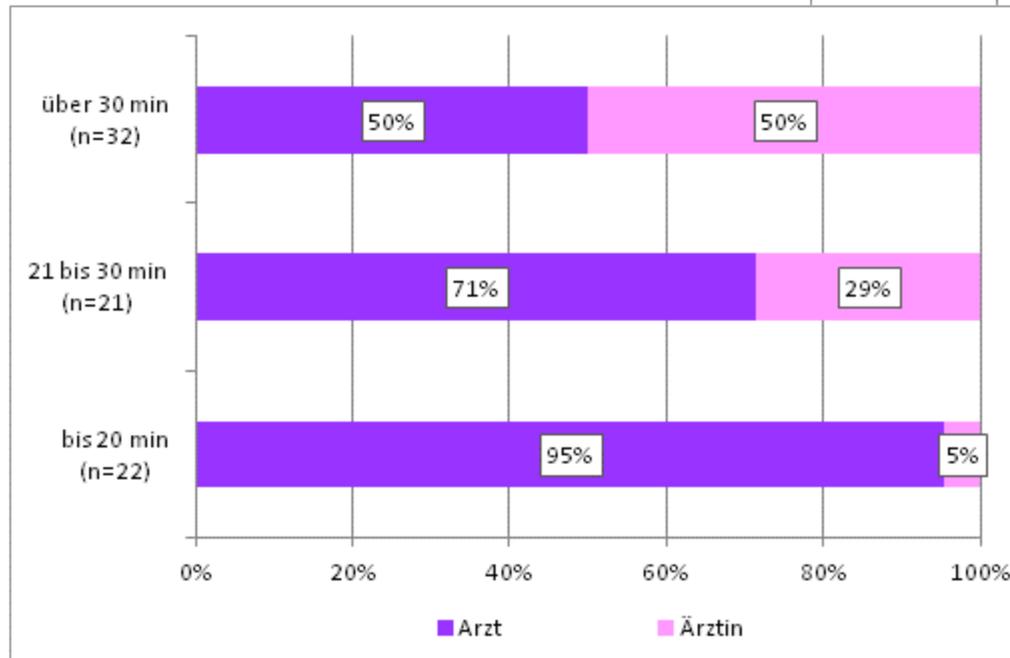
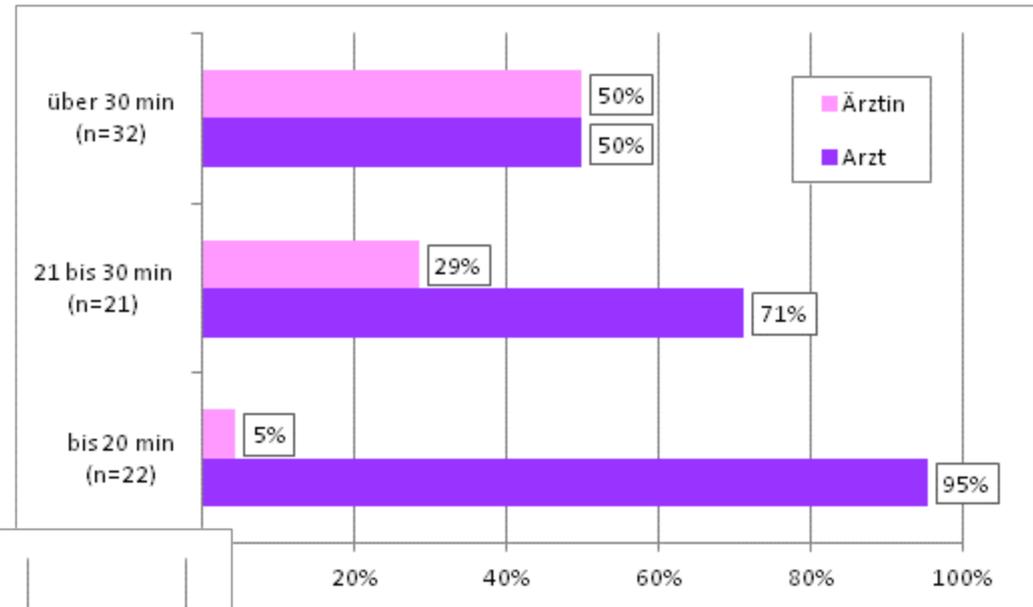
Gruppiertes Balkendiagramm



Gestapeltes Balkendiagramm

Gesprächsdauer nach Geschlecht der/des Ärzt:in → Dauer ist 100%

**Gruppiertes
Balkendiagramm**



**Gestapeltes
Balkendiagramm**

Signifikanztests:

**Chi-Quadrat-Test
= Häufigkeitsvergleich
bei Kreuztabellen**

Beobachtete Anzahl	Anzahl der tatsächlich beobachteten Fälle in einer Zelle
Zeilensumme	einfache Häufigkeits- od. Randverteilung der Zeilenvariable
Spaltensumme	einfache Häufigkeits- od. Randverteilung der Spaltenvariable
% von Zeilenvariable	Zeilenweise Prozentuierung: Basis der Prozentwerte ist die Zeilensumme
% von Spaltenvariable	Spaltenweise Prozentuierung: Basis der Prozentwerte ist die Spaltensumme
% von Gesamtzahl	Gesamtprozentwerte: Fälle in den Zellen werden als Prozentanteile an allen gültigen Fällen ausgedrückt
Erwartete Anzahl	Erwartete Werte: Anzahl der Fälle in einer Zelle, die erwartet werden, wenn kein Zusammenhang zw. den beiden Variablen besteht bzw. wenn sie unabhängig voneinander sind
Residuen	Differenzen zwischen den beobachteten und erwarteten Werten
Standardisierte Residuen	Standardisierte Residuen: Diese Differenzen werden an den erwarteten Werten standardisiert
Chi-Quadrat-Testwert	Testverfahren für die Analyse von Häufigkeitsunterschieden bestimmter Merkmale durch den Vergleich von beobachteten und erwarteten Häufigkeiten
Signifikanz	Irrtums-WS wird auf Basis des Chi-Quadrat-Wertes und der Freiheitsgrade (df) ermittelt <i>größer als α</i> : Nullhypothese (es besteht kein Zusammenhang) wird vorläufig beibehalten. <i>kleiner als α</i> : Nullhypothese kann abgelehnt und die Alternativhypothese (es besteht ein Zusammenhang) angenommen werden.
Anz der Zellen (%) mit erwarteter Häufigkeit <5	Der angeführte Anteil sollte 20 % nicht überschreiten

**Unterscheidet sich die Wortanzahl der Ärzt:innen
signifikant nach deren Geschlecht?**

ODER: Sprechen Ärztinnen signifikant mehr Wörter als Ärzte?

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A Wortanzahl ÄrztIn	bis zu 1900 Wörter	Anzahl	16	2	18
		%	61,5%	18,2%	48,6%
	mehr als 1901 Wörter	Anzahl	10	9	19
		%	38,5%	81,8%	51,4%
Gesamt		Anzahl	26	11	37
		%	100,0%	100,0%	100,0%

Erwartete Anzahl bei Unabhängigkeit

= Wenn Ärzte und Ärztinnen gleich viel Wörter sprechen.

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A Wortanzahl ÄrztIn	bis zu 1900 Wörter	Anzahl	16	2	18
		Erwartete Anzahl	12,6	5,4	18,0
	mehr als 1901 Wörter	Anzahl	10	9	19
		Erwartete Anzahl	13,4	5,6	19,0
Gesamt		Anzahl	26	11	37
		Erwartete Anzahl	26,0	11,0	37,0

Randsummen werden als gegeben angenommen:

Erwartete Anzahl in den einzelnen Zellen =

(Randsumme Zeile x Randsumme Spalte) / Gesamtsumme

Bsp.: Arzt bis zu 1900 Wörter: $(18 \times 26) / 37 = 12,6$

Absolutes Residuum = Differenz zwischen tatsächlicher und erwarteter Häufigkeit
 = Abweichung tatsächlicher Häufigkeit zur Unabhängigkeit (=Zufall)

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A Wortanzahl ÄrztIn	bis zu 1900 Wörter	Anzahl	16	2	18
		Erwartete Anzahl	12,6	5,4	18,0
		Residuen	3,4	-3,4	
	mehr als 1901 Wörter	Anzahl	10	9	19
		Erwartete Anzahl	13,4	5,6	19,0
		Residuen	-3,4	3,4	
Gesamt		Anzahl	26	11	37
		Erwartete Anzahl	26,0	11,0	37,0

Residuum = Tatsächliche Zellhäufigkeit – erwartete Zellhäufigkeit

Bsp.: Arzt bis zu 1900 Wörter: $16 - 12,6 = 3,4$

Standardisiertes Residuum = Standardisiertes absolutes Residuum
 = standardisierte Abweichung tatsächlicher Hfkt. zur Unabhängigkeit (=Zufall)

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A Wortanzahl ÄrztIn	bis zu 1900 Wörter	Anzahl	16	2	18
		Erwartete Anzahl	12,6	5,4	18,0
	Residuen	3,4	-3,4		
	Standardisierte Residuen	0,9	-1,4		
	mehr als 1901 Wörter	Anzahl	10	9	19
		Erwartete Anzahl	13,4	5,6	19,0
	Residuen	-3,4	3,4		
	Standardisierte Residuen	-0,9	1,4		
Gesamt		Anzahl	26	11	37
		Erwartete Anzahl	26,0	11,0	37,0

$$\text{Standardisiertes Residuum} = \frac{\text{abs. Res}}{\sqrt{\text{erw. Hfkt}}}$$

Bsp.: Arzt bis zu 1900 Wörter : $3,4 / \sqrt{12,6} = 0,9$

Kreuztabelle mit Gewichtung x 5

= die Gesamtstichprobe von 37 untersuchten Gesprächen wird ver-5-facht.

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A Wortanzahl ÄrztIn	bis zu 1900 Wörter	Anzahl	80	10	90
		Erw artete Anzahl	63,2	26,8	90,0
	Residuen	16,8	-16,8		
	Standardisierte Residuen	2,1	-3,2		
	mehr als 1901 Wörter	Anzahl	50	45	95
		Erw artete Anzahl	66,8	28,2	95,0
	Residuen	-16,8	16,8		
	Standardisierte Residuen	-2,1	3,2		
Gesamt		Anzahl	130	55	185
		Erw artete Anzahl	130,0	55,0	185,0

Kreuztabelle mit Gewichtung x 10

= die Gesamtstichprobe von 37 untersuchten Gesprächen wird ver-10-facht.

wort_A Wortanzahl ÄrztIn * sex_A Geschlecht ÄrztIn Kreuztabelle

			sex_A Geschlecht ÄrztIn		Gesamt
			0 männlich	1 weiblich	
wort_A Wortanzahl ÄrztIn	bis zu 1900 Wörter	Anzahl	160	20	180
		Erwartete Anzahl	126,5	53,5	180,0
	Residuen	33,5	-33,5		
	Standardisierte Residuen	3,0	-4,6		
	mehr als 1901 Wörter	Anzahl	100	90	190
		Erwartete Anzahl	133,5	56,5	190,0
	Residuen	-33,5	33,5		
	Standardisierte Residuen	-2,9	4,5		
Gesamt	Anzahl		260	110	370
	Erwartete Anzahl		260,0	110,0	370,0

Wichtig bei Signifikanztests bei Kreuztabellen: Voraussetzung:

Die erwartete Anzahl darf in keiner Zelle weniger als 5 sein.

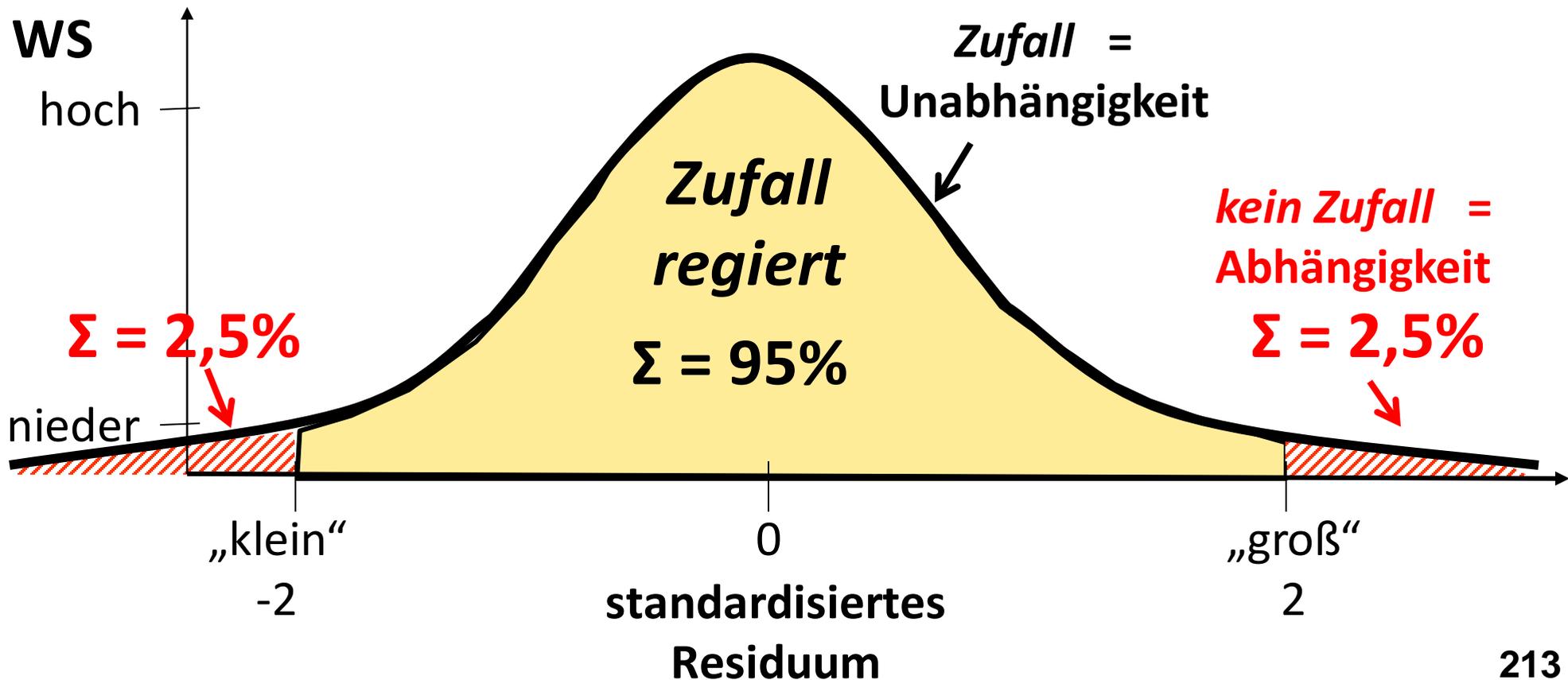
Weniger streng: Die erwartete Anzahl darf in 20% der Zellen weniger als 5 sein.

Unser Bsp.: auch bei den ungewichteten Daten wird diese Voraussetzung erfüllt.

Theoretische Normalverteilung (= Wahrscheinlichkeitsverteilung)

Anwendung z.B. als Rechenschablone: Standardnormalverteilung

StdNV: ist normiert: Mittelwert = 0, StdAbw. = 1



Gedankenexperiment:

Summieren Sie die absoluten Residuen einer Kreuztabelle.

Was kommt heraus?

Summieren Sie die standardisierten Residuen einer Kreuztabelle.

Was kommt heraus?

= ???

$$\sum \frac{f_o - f_e}{\sqrt{f_e}} \quad \text{Prüfmaß ?} = 0$$

Quadrieren:

$$\sum \frac{(f_o - f_e)^2}{f_e} = \chi^2$$

Quadrierte Summe aller Abweichungen (=Residuen)

**Abweichung von erwarteter Verteilung
(oder: Zufallsverteilung)**

**Große Abweichung vom Zufall
= hohes Chi-Quadrat**

Test bei Kreuztabellen: Prüfmaß Chi-Quadrat und Prüfmaß Cramers´ s V

Chi-Quadrat = Prüfmaß zur Testung eines signifikanten Unterschiedes

Dieses Prüfmaß kann theoretisch zwischen 0 und unendlich liegen.
Ist nur interpretierbar in Kombination mit der Signifikanz.
Sagt aus, ob ein signifikanter Unterschied zwischen den beiden Merkmalen der Kreuztabelle besteht – oder nicht.

Cramer´ s V = Prüfmaß zur Testung der Stärke eines signifikanten Zusammenhangs

Diese „Stärke“ ist normiert zwischen 0 und 1.
Verhältnis des tatsächlichen Chi-Quadrat-Wertes zum maximal möglichen Chi-Quadrat-Wert (bei gleich Randsummen).
Ist nur bedeutsam, wenn ein signifikanter Unterschied besteht.
Von erwähnenswerten Zusammenhängen spricht man ab $V = \sim 0,3$.

Maßzahlen für die Stärke des Zusammenhangs

(bei Kreuztabellen auf Basis von Chi-Quadrat)

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

Phi-Koeffizient:

Werte zwischen 0 und 1
(nur für 2x2 Kreuztabellen)

Cramer's V-Koeffizient:

Werte zwischen 0 und 1
(kann als % interpretiert werden)

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$

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

Kontingenz-Koeffizient:

Werte zwischen 0 und (fast) 1

Signifikanz

Jedes Prüfmaß hat eine berechenbare Signifikanz.

Signifikanz

= Wahrscheinlichkeit des jeweiligen Prüfmaß unter Voraussetzung von Unabhängigkeit (= Zufall) (= kein Unterschied) (= kein Zusammenhang)

Unabhängigkeit = Zufall = kein Unterschied = keine Gesetzmäßigkeit

Abhängigkeit = kein Zufall = Unterschied / Zusammenhang = Gesetzmäßigkeit

Die Signifikanz des Prüfmaß spricht
entweder für Unabhängigkeit (H_0) oder für Abhängigkeit (H_1).

Arten von Hypothesen

Nullhypothese (H_0): Ausgangspunkt der Testung:

→ Es besteht **kein** signifikanter Unterschied/Zusammenhang

*„Ärzte und Ärztinnen weisen in den Gesprächen
die **gleiche Wortanzahl** auf.“*

Die tatsächlichen Zellhäufigkeiten der Kreuztabelle entsprechen in etwa den erwarteten Zellhäufigkeiten.

Alternativhypothese (H_1): Hypothese zum Zusammenhang:

→ Es besteht **ein** signifikanter Unterschied/Zusammenhang

*„Ärzte und Ärztinnen weisen in den Gesprächen
eine **unterschiedliche Wortanzahl** auf.“*

Die tatsächlichen Zellhäufigkeiten der Kreuztabelle weichen signifikant von den erwarteten Zellhäufigkeiten ab.

Regeln bei der Hypothesentestung – Konvention

Nullhypothese:

→ wenn der Signifikanzwert groß ist ($p > 0,05$) dann wird die Nullhypothese beibehalten (kein signifikanter Unterschied).

Alternativhypothese: Hypothese zum Zusammenhang:

→ wenn der Signifikanzwert sehr klein ist ($p \leq 0,05$) dann wird die Alternativhypothese angenommen (signifikanter Unterschied/Zusammenhang)

Die **Fehlerwahrscheinlichkeit** ist dann kleiner als 0,05 (oder: 5%).

Das Risiko, einen falschen Zusammenhang anzunehmen, liegt unter 5%.

Chi-Quadrat-Test:

Vergleich zwischen beobachteten und erwarteten Häufigkeiten

→ **Voraussetzungen:**

- Kreuztabelle mit zwei Variablen mit kategorialem Datenniveau
- ausreichende Fallzahlen in den Zellen der Kreuztabelle

→ **Nullhypothese:**

Ärzte und ÄrztInnen sprechen gleich viele Wörter.



→ **Rechenlogik:** Untersucht wird die Abweichung zwischen den beobachteten und den erwarteten Häufigkeiten der Kreuztabelle.

→ Prüfmaß = Chi-Quadrat

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	5,816 ^b	1	,016
Anzahl der gültigen Fälle	37		

Prüfmaß Chi-Quadrat

Signifikanz
 (= WS unter der Voraussetzung von Unabhängigkeit)
Fehler-WS = 1,6%

- a. Wird nur für eine 2x2-Tabelle berechnet
- b. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 5,35.

Es zeigt sich ein signifikanter Unterschied hinsichtlich der Anzahl der gesprochenen Wörter nach dem Geschlecht der Ärzt:innen ($\text{Chi}^2=5,816$, $\text{df}=1$, $p=0,016$).

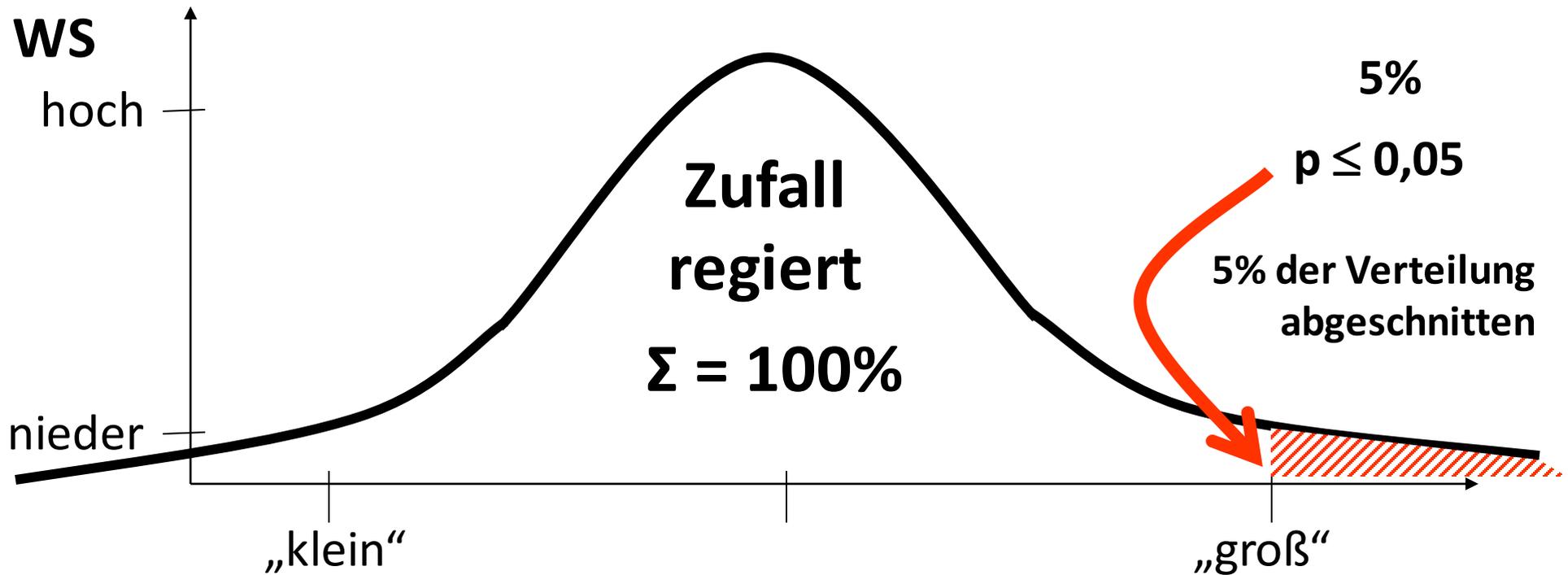
Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,396	,016
	Cramer-V	,396	,016
Anzahl der gültigen Fälle		37	

Stärke des Zusammenhangs wenn $> 0,3$ stark

- a. Die Null-Hypothese wird nicht angenommen.

Wahrscheinlichkeitsverteilung von einem Prüfmaß



Der Wert, ab welchem das Prüfmaß signifikant ist, variiert je nach Fallzahl der Stichprobe.
→ Je größer die Stichprobe, desto kleiner kann dieser Wert sein.

Fallzahl	Prüfmaß Chi-Quadrat	Prüfmaß Cramer's V	Signifikanz
n = 37	$\chi^2 = 5,8$	V = 0,396	p = 0,016
n = 185	$\chi^2 = 29,1$	V = 0,396	p = 0,000
n = 370	$\chi^2 = 58,2$	V = 0,396	p = 0,000

Für die Hypothese
**„Ärzte und Ärztinnen weisen in den Gesprächen
eine unterschiedliche Wortanzahl auf.“**
ergibt sich bereits ab einer Fallzahl von 37 Gesprächen
ein **signifikanter Unterschied**:
Ärztinnen sprechen signifikant mehr Wörter als Ärzte.

Beispiel Anzahl der gesprochenen Wörter der Patient:innen nach Geschlecht der/des Patient:in

wort_P Wortanzahl PatientIn * sex_P Geschlecht PatientIn Kreuztabelle

			sex_P Geschlecht PatientIn		Gesamt
			0 männlich	1 weiblich	
wort_P	bis 750 Wörter	Anzahl	6	12	18
Wortanzahl PatientIn		%	54,5%	46,2%	48,6%
	mehr als 751 Wörter	Anzahl	5	14	19
		%	45,5%	53,8%	51,4%
Gesamt		Anzahl	11	26	37
		%	100,0%	100,0%	100,0%

Beispiel Anzahl der gesprochenen Wörter der Patient:innen nach Geschlecht der/des Patient:in

wort_P Wortanzahl PatientIn * sex_P Geschlecht PatientIn Kreuztabelle

			sex_P Geschlecht PatientIn		Gesamt
			0 männlich	1 weiblich	
wort_P Wortanzahl PatientIn	bis 750 Wörter	Anzahl	6	12	18
		Erwartete Anzahl	5,4	12,6	18,0
		Residuen	,6	-,6	
		Standardisierte Residuen	,3	-,2	
	mehr als 751 Wörter	Anzahl	5	14	19
		Erwartete Anzahl	5,6	13,4	19,0
		Residuen	-,6	,6	
		Standardisierte Residuen	-,3	,2	
Gesamt	Anzahl	11	26	37	
	Erwartete Anzahl	11,0	26,0	37,0	

Chi-Quadrat-Tests

	Wert	df	Asymmetrisch Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	,218 ^b	1	,641
Anzahl der gültigen Fälle	37		

Prüfmaß Chi-Quadrat

Signifikanz
(= WS unter der
Voraussetzung von
Unabhängigkeit)
Fehler-WS = 64%

- a. Wird nur für eine 2x2-Tabelle berechnet
- b. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 5,35.

Es zeigt sich kein signifikanter Unterschied hinsichtlich der Anzahl der gesprochenen Wörter nach dem Geschlecht der Patient:innen ($\text{Chi}^2=0,218$, $\text{df}=1$, $p=0,641$).

Symmetrische Maße

		Wert	Näherung sw eise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,077	,641
	Cramer-V	,077	,641
Anzahl der gültigen Fälle		37	

Stärke des
Zusammenhangs
wenn $< 0,2$ schwach

- a. Die Null-Hypothese wird nicht angenommen.

Fallzahl	Prüfmaß Chi-Quadrat	Prüfmaß Cramer's V	Signifikanz
n = 37	$\chi^2 = 0,2$	V = 0,077	p = 0,641
n = 185	$\chi^2 = 1,1$	V = 0,077	p = 0,297
n = 370	$\chi^2 = 2,2$	V = 0,077	p = 0,140
n = 740	$\chi^2 = 4,3$	V = 0,077	p = 0,037
n = 1110	$\chi^2 = 6,5$	V = 0,077	p = 0,011
n = 1850	$\chi^2 = 10,9$	V = 0,077	p = 0,001

Erst ab einer Fallzahl von rund $n \sim 700$ Gesprächen zeigt unsere Kreuztabelle einen **signifikanten Unterschied** hinsichtlich der Wortanzahl nach Geschlecht der/der Pat.:
 Erst wenn über 700 Gespräche untersucht werden WÜRDEN (Simulation!),
 wird dieser geringe Unterschied signifikant.

Statistische Verfahren – Auswahl des Testverfahrens

Je nach Fragestellung und Datenniveau wird das passende statistische Verfahren zur Analyse ausgewählt.

Bei der Testwahl ist entscheidend, dass das Verfahren mit der „höheren **Teststärke**“ gewählt wird.

Bei metrischen Daten sollten daher **parametrische Tests** angewandt werden. Um dies sicherzustellen, wird die metrische Datenqualität meist auf Voraussetzungen (beispielsweise Normalverteilung) überprüft.

Bei ordinalen Daten, metrischen Daten in geringen Fallzahlen die nicht normalverteilt sind, werden **parameterfreie Testverfahren** angewandt.

Aufstellen von **Nullhypothese** und **Alternativhypothese**:

Nullhypothese: Es gibt keinen Gruppenunterschied / keinen Zusammenhang.

Alternativhypothese: Es gibt einen Gruppenunterschied / Zusammenhang.

Statistische Verfahren – Hypothesenentscheidung

Das Ergebnis eines Testverfahrens besteht aus einem oder mehreren Testwerten (=Prüfmaß) und deren Signifikanz.

Die Signifikanz eines Testergebnisses wird interpretiert: Sie sagt aus, ob das Ergebnis zufällig zustande gekommen ist, oder ob die Gesetzmäßigkeit (H_A) bestätigt werden konnte.

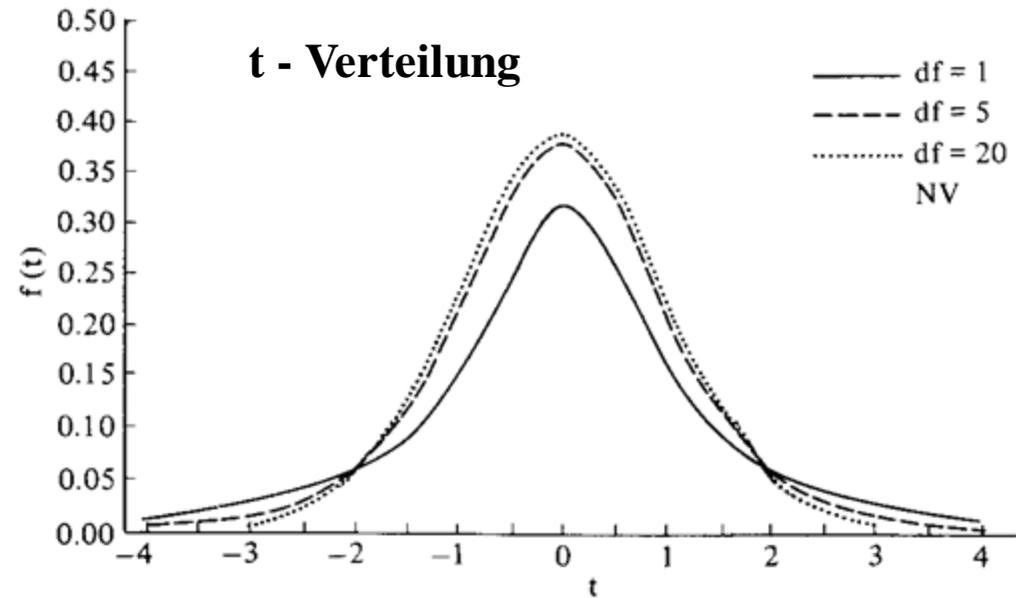
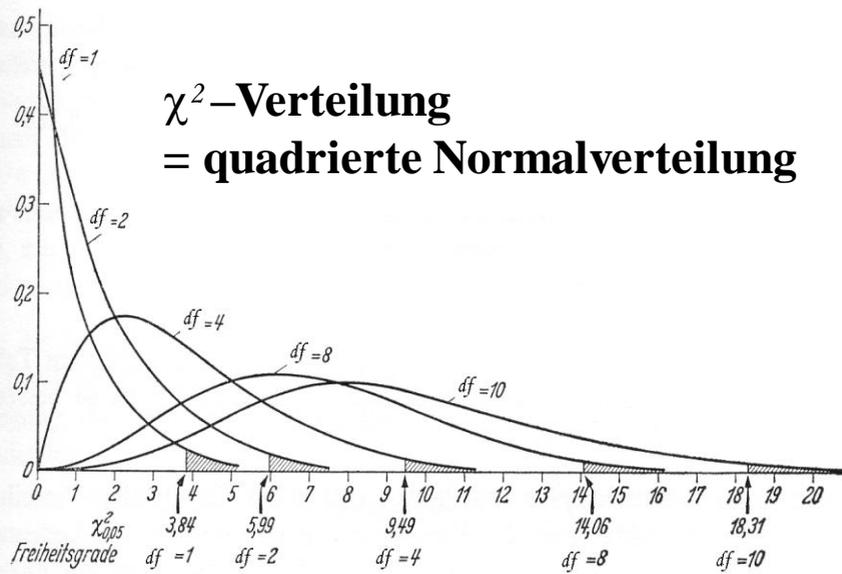
Bei einigen Verfahren (Korrelation, Regression) sind die Testwerte bereits zwischen festen Grenzen normiert und können daher auch inhaltlich interpretiert werden (z.B. Stärke einer Korrelation zwischen 0 und 1).

Die Signifikanz des Testwertes gibt an, wie hoch die WS ist, dass das Stichprobenergebnis aus einer GG stammt, in der die Nullhypothese gilt.

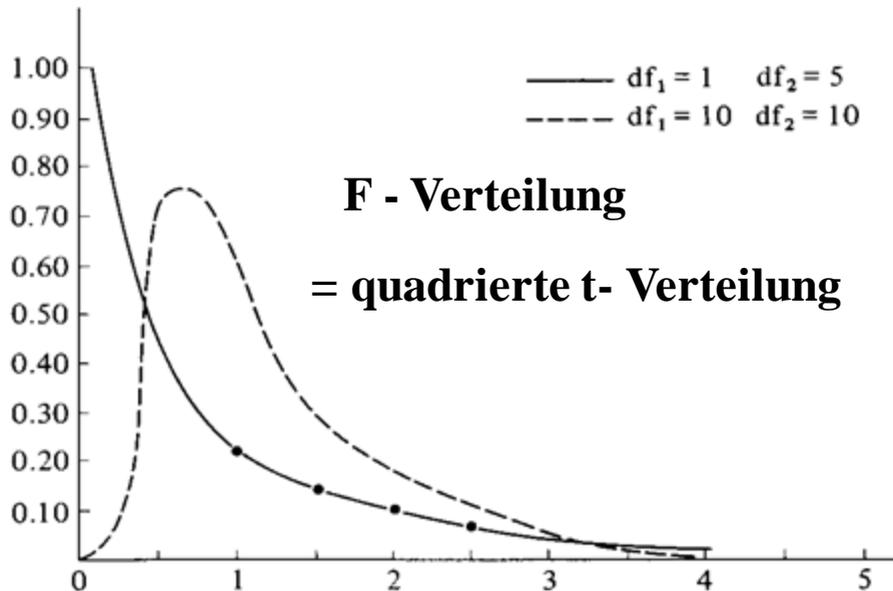
Ist diese WS > 0,05 (oder 5%) wird die Nullhypothese beibehalten.

Ist diese WS ≤ 0,05 (oder 5%) wird die Alternativhypothese angenommen.

Alternativhypothese bedeutet: Es ist sehr unwahrscheinlich, dass in der GG die Nullhypothese gilt, weshalb davon ausgegangen wird, dass die Gesetzmäßigkeit der Stichprobe gilt.



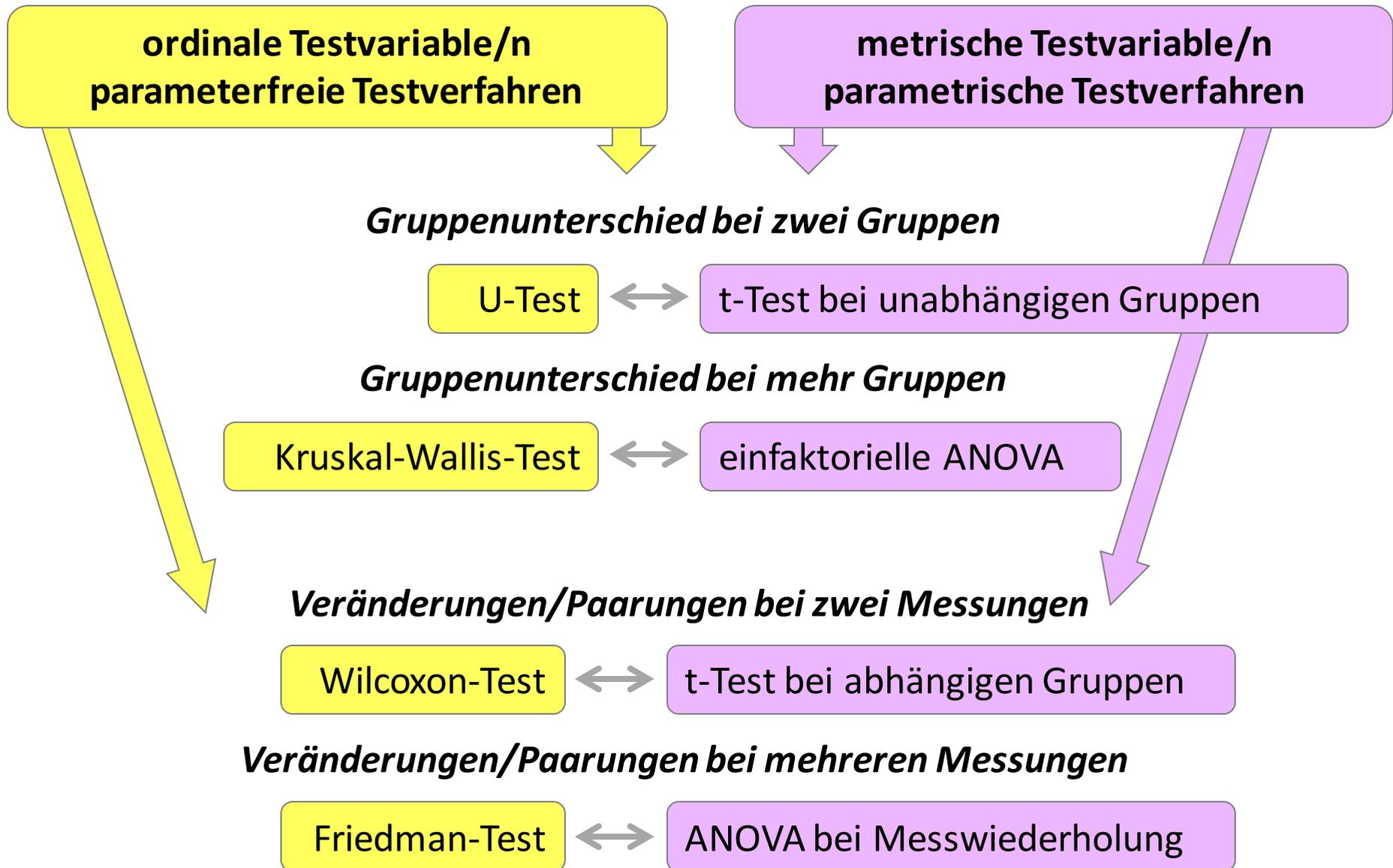
Wahrscheinlichkeitsverteilungen einiger Prüfmaße



Normalverteilung

z - Verteilung

Überblick über die wichtigsten Signifikanztests



Testen von Unterschiedshypothesen:

Parameterfreie Tests

U-Test und Kruskal-Wallis-Test
Wilcoxon-Test und Friedman-Test

U-Test: Vergleich der Verteilungen zwischen **zwei** Gruppen

→ Voraussetzungen:

- mind. **ordinales** Datenniveau bei der Test-Variable (abhängige Variable)
- zwei Ausprägungen bei der Gruppenvariable (unabhängige Variable)

→ Nullhypothese:

Die Wichtigkeit von Grammatik bei Simultanübersetzungen unterscheidet sich nicht zwischen **zwei** Berufsgruppen.



→ **Rechenlogik:** Alle Werte beider Gruppen werden in eine gemeinsame Reihenfolge gebracht, die Rangplätze werden ermittelt, die mittleren Ränge beider Gruppen werden verglichen.

→ Prüfmaß = z

Beispiel: Ordinale Testvariable und eine Gruppenvariable

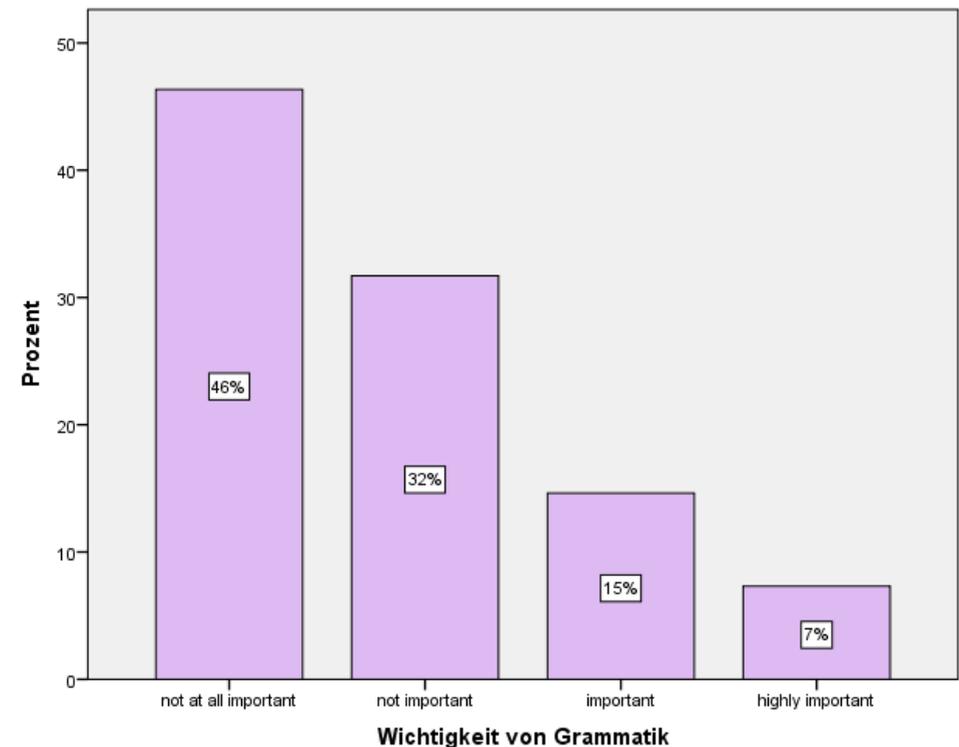
Untersucht wird, ob sich Berufsgruppen hinsichtlich der Wichtigkeit von korrekter Grammatik bei Simultanübersetzungen unterscheiden.

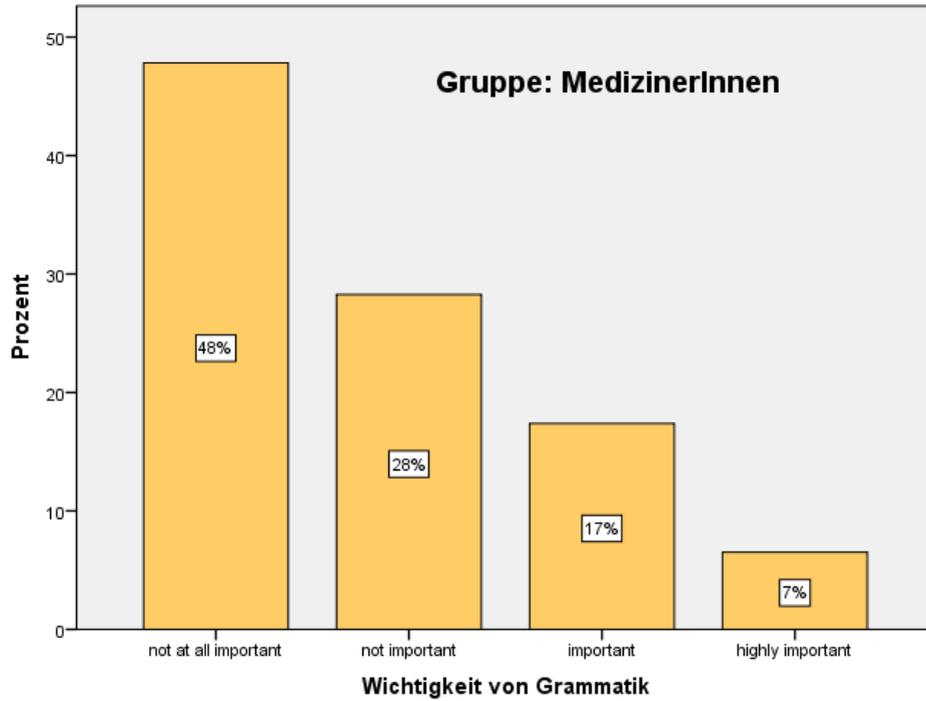
group Gruppe

		Häufigkeit
Gültig	TechnikerInnen	58
	MedizinerInnen	92
	DiplomatInnen	96
	Gesamt	246

bew Grammatik

		Häufigkeit	Prozent	Gültige Prozente
Gültig	not at all important	114	46,3	46,3
	not important	78	31,7	31,7
	important	36	14,6	14,6
	highly important	18	7,3	7,3
	Gesamt	246	100,0	100,0





bew Grammatik^a

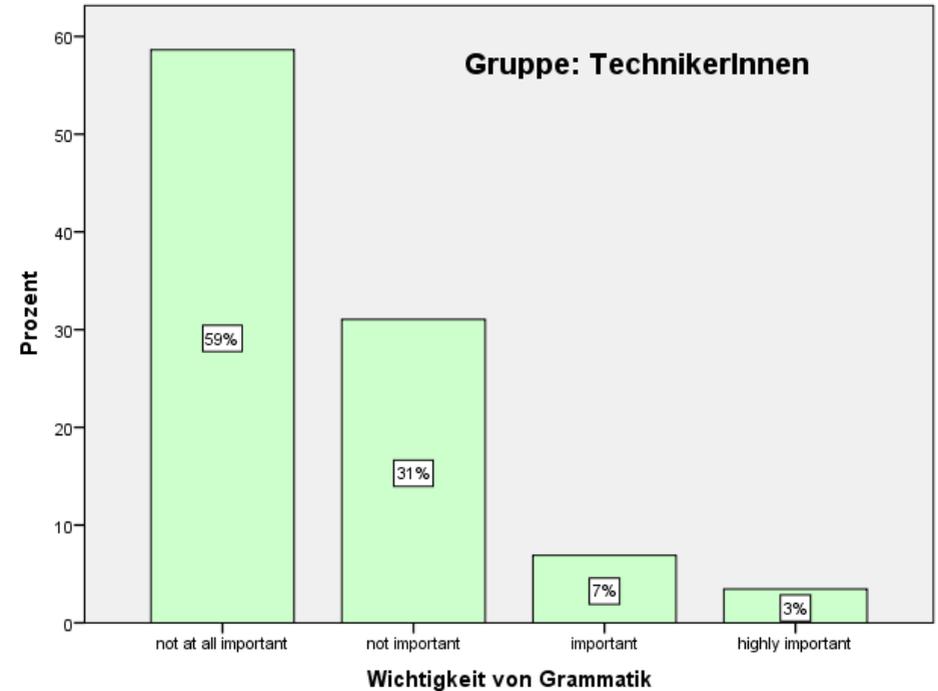
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	not at all important	44	47,8	47,8	47,8
	not important	26	28,3	28,3	76,1
	important	16	17,4	17,4	93,5
	highly important	6	6,5	6,5	100,0
Gesamt		92	100,0	100,0	

a. group Gruppe = MedizinerInnen

bew Grammatik^a

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	not at all important	34	58,6	58,6	58,6
	not important	18	31,0	31,0	89,7
	important	4	6,9	6,9	96,6
	highly important	2	3,4	3,4	100,0
Gesamt		58	100,0	100,0	

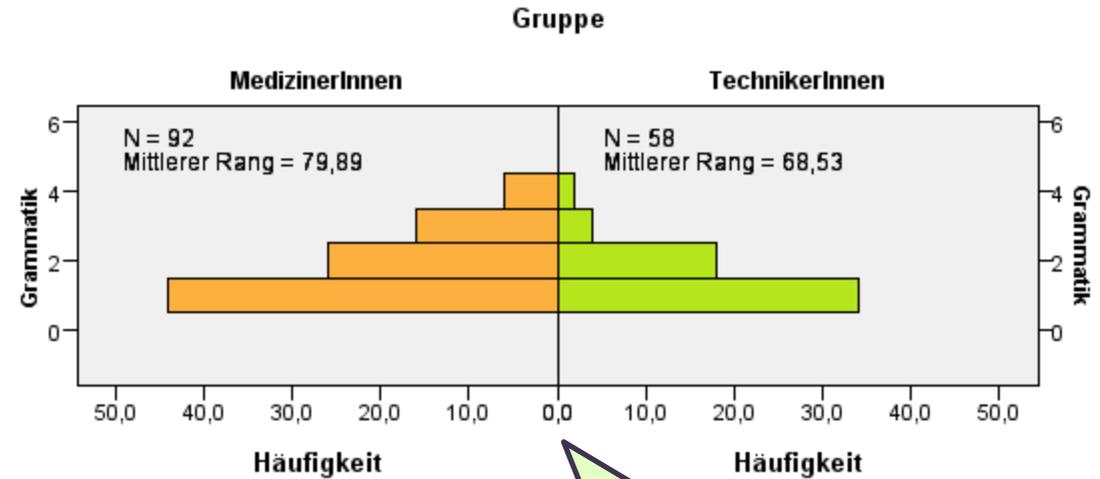
a. group Gruppe = TechnikerInnen



Mann-Whitney-U-Test bei unabhängigen Stichproben

Mann-Whitney-U-Test

Gesamtanzahl	150
Mann-Whitney-U	3.072,000
Wilcoxon-W	7.350,000
Teststatistik	3.072,000
Standardfehler	236,309
Standardisierte Teststatistik	1,710
Asymptotische Sig. (2-seitiger Test)	,087



Prüfmaß z

Mittlere Ränge
von 1 bis 150 Rangplätzen

WS von Prüfmaß z unter Voraussetzung
der Nullhypothese = 8,7%

Bei einer Fehlerwahrscheinlichkeit von $> 5\%$ wird die **Nullhypothese (H_0) beibehalten**:
Es besteht kein signifikanter Unterschied zwischen den beiden Berufsgruppen hinsichtlich der Wichtigkeit von korrekter Grammatik bei Simultanübersetzungen ($z=1,710$, $n=150$, $p=0,087$).

Kruskal-Wallis-Test: Vergleich der Verteilungen zwischen **mehreren** Gruppen

→ Voraussetzungen:

- mind. **ordinales** Datenniveau bei der Test-Variable (abhängige Variable)
- mehrere Ausprägungen bei der Gruppenvariable (unabhängige Variable)

→ Nullhypothese:

Die Wichtigkeit von Grammatik bei Simultanübersetzungen unterscheidet sich nicht zwischen **drei** Berufsgruppen.

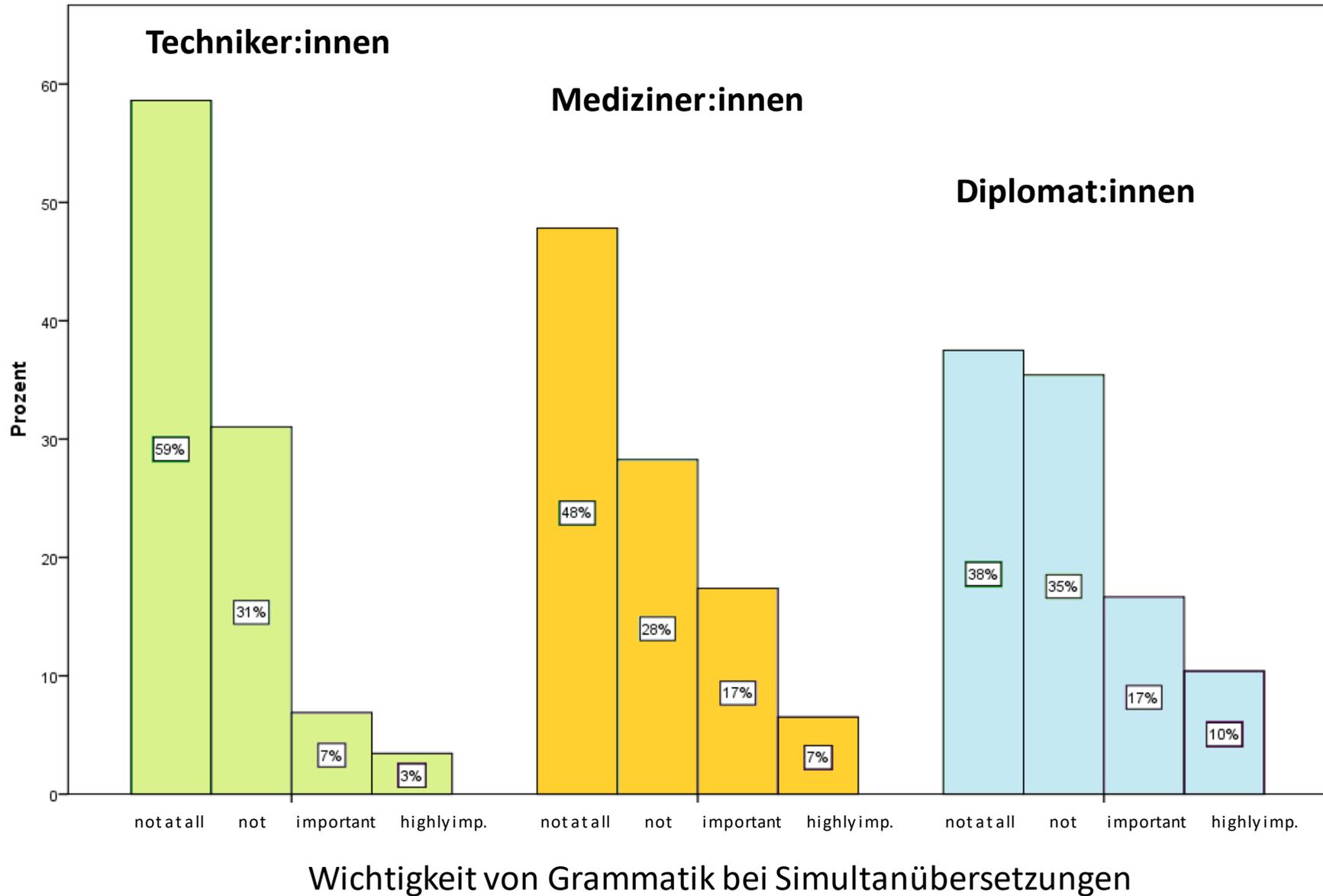
**Berufsgruppe
(unabhängig)**

hat Einfluss auf

**Wichtigkeit von Grammatik
(abhängig)**

→ **Rechenlogik:** Alle Werte beider Gruppen werden in eine gemeinsame Reihenfolge gebracht, die Rangplätze werden ermittelt, die mittleren Ränge beider Gruppen werden verglichen.

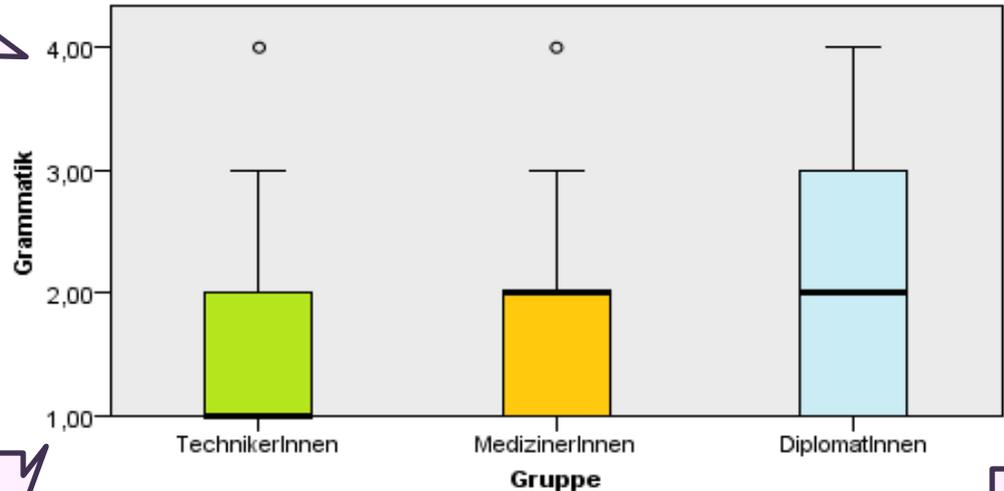
→ Prüfmaß = **H**



Kruskal-Wallis-Test

4 = highly imp.

Kruskal-Wallis-Test bei unabhängigen Stichproben



1 = not at all imp.

Gesamtanzahl	246
Teststatistik	8,303
Freiheitsgrade	2
Asymptotische Sig. (2-seitiger Test)	,016

Prüfmaß

WS von Prüfmaß z unter Voraussetzung der Nullhypothese

Bei einer Fehler-WS von $< 5\%$ wird die **Alternativhypothese (H_A)** angenommen:

Es besteht ein signifikanter Unterschied zwischen den drei Berufsgruppen hinsichtlich der Wichtigkeit von korrekter Grammatik bei Simultanübersetzungen ($H=8,303$, $n=246$, $p=0,016$).

Anhand der Boxplots / Diagramme wird ersichtlich, dass die Berufsgruppe der DiplomatInnen korrekte Grammatik wichtiger einstuft, als die beiden anderen Berufsgruppen. (unwichtig = 1, sehr wichtig = 4)

Wilcoxon-Test: Vergleich der Rangplätze einer Veränderung zwischen **zwei** Messzeitpunkten

→ Voraussetzungen:

- zwei Testvariablen, die inhaltlich voneinander abhängig sind
- mindestens **ordinales** Datenniveau der Test-Variablen

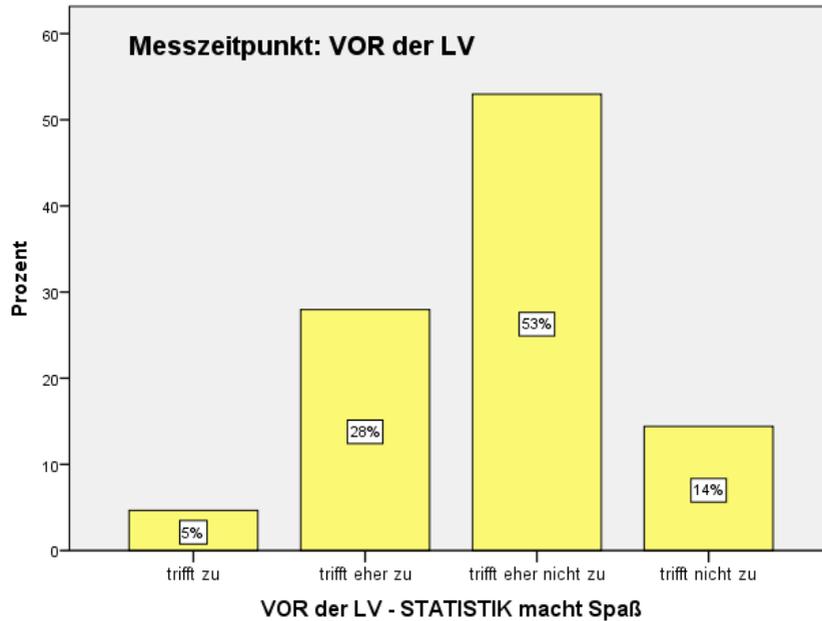
→ Nullhypothese:

Die Meinung, dass Statistik Spaß macht, verändert sich nicht zwischen Beginn und Ende der Statistik-Vorlesung.

- ### → Rechenlogik: Testen der Differenz zwischen VORHER und NACHHER:
- negative Ränge: Nach der VO nimmt die Meinung, dass Statistik Spaß macht, ab
positive Ränge: Nach der VO nimmt die Meinung, dass Statistik Spaß macht, zu
Bindungen: Die Meinung, dass Statistik Spaß macht, bleibt gleich

- ### → Prüfmaß = z oder Wilcoxon-W

Beispiel: Ordinale Testvariable zu zwei Messzeitpunkten

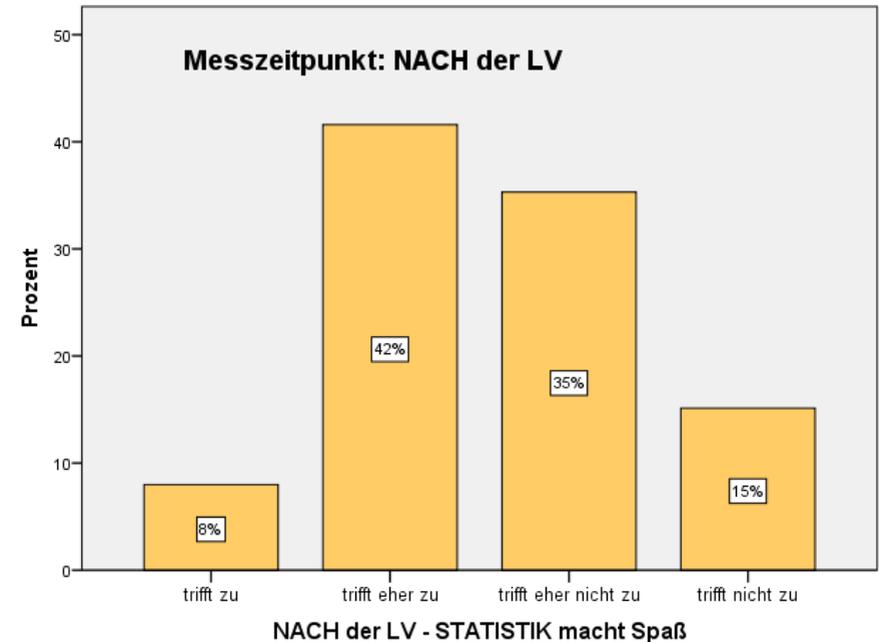


spass_vorher VOR der LV - STATISTIK macht Spaß

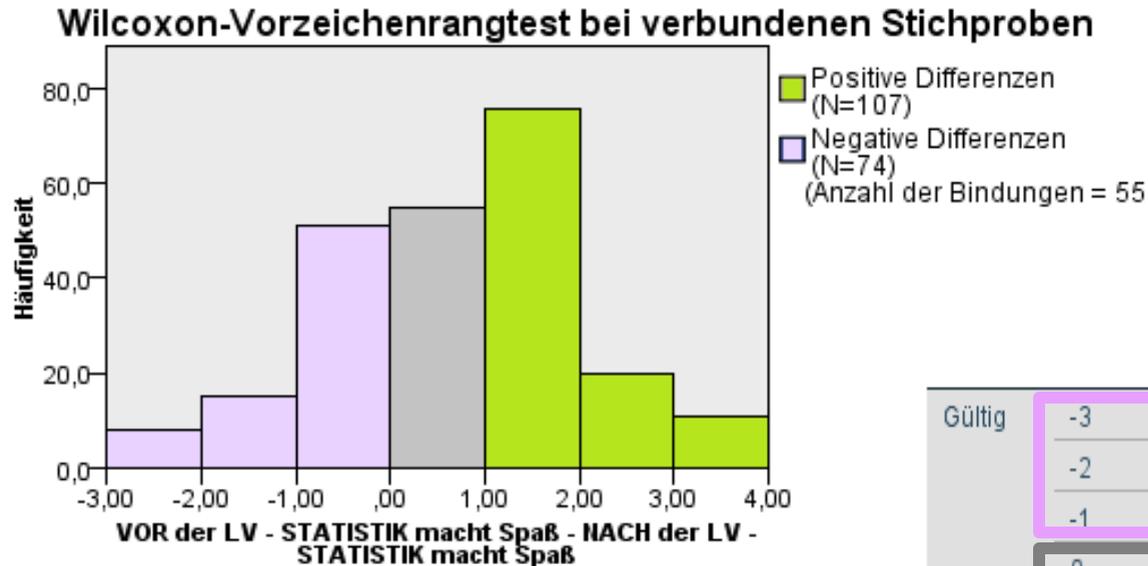
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft zu	11	4,5	4,7	4,7
	trifft eher zu	66	27,2	28,0	32,6
	trifft eher nicht zu	125	51,4	53,0	85,6
	trifft nicht zu	34	14,0	14,4	100,0
	Gesamt	236	97,1	100,0	
Fehlend	System	7	2,9		
Gesamt		243	100,0		

spass_nachher NACH der LV - STATISTIK macht Spaß

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft zu	19	7,8	8,0	8,0
	trifft eher zu	99	40,7	41,6	49,6
	trifft eher nicht zu	84	34,6	35,3	84,9
	trifft nicht zu	36	14,8	15,1	100,0
	Gesamt	238	97,9	100,0	
Fehlend	System	5	2,1		
Gesamt		243	100,0		



Differenz = Veränderung zwischen VORHER und NACHHER



Differenz = Wert „vorher“
minus Wert „nachher“

differenz

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	-3	8	3,3	3,4	3,4
	-2	15	6,2	6,4	9,7
	-1	51	21,0	21,6	31,4
	0	55	22,6	23,3	54,7
	1	76	31,3	32,2	86,9
	2	23	9,2	8,5	95,3
	3	11	4,5	4,7	100,0
	Gesamt	236	97,1	100,0	
Fehlend	System	7	2,9		
	Gesamt	243	100,0		

unverändert

nach der LV macht
Statistik **weniger** Spass

nach der LV macht
Statistik **mehr** Spass

Wilcoxon-Test

Gesamtanzahl	236
Teststatistik	9.656,000
Standardfehler	674,196
Standardisierte Teststatistik	2,107
Asymptotische Sig. (2-seitiger Test)	,035

Prüfmaß z

WS von Prüfmaß z unter Voraussetzung der Nullhypothese

Bei einer Fehler-WS von 3,5% wird die **Alternativhypothese (H_1) angenommen:**

Die Meinung, dass Statistik Spaß macht, verändert sich signifikant zwischen Beginn und Ende der Vorlesung ($z=2,107$, $n=236$, $p=0,035$).

Anhand der Verteilung der Differenzen wird deutlich, dass die Meinung, dass Statistik Spass macht, zwischen Beginn und Ende der Vorlesung signifikant zunimmt.

Friedman-Test:

Vergleich der Rangplätze einer Veränderung bei **mehreren** Zeitpunkten

→ Voraussetzungen:

- **ordinales** Datenniveau der Test-Variablen
- mehrere Testvariablen, die inhaltlich voneinander abhängig sind

→ Nullhypothese:

Die Meinung, dass Statistik Spaß macht, verändert sich nicht zwischen Beginn der Vorlesung, Ende der Vorlesung und Ende der Übung.

→ **Rechenlogik:** Testen der Differenz zwischen den Messzeitpunkten
Signifikant: Wenn Unterschied zumindest zwischen zwei Messzeitpunkten

→ Prüfmaß = Friedman-Chi-Quadrat

Beispiel: Ordinale Testvariable zu mehreren Messzeitpunkten

spass_vorher VOR der LV - STATISTIK macht Spaß

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft zu	11	4,5	4,7	4,7
	trifft eher zu	66	27,2	28,0	32,6
	trifft eher nicht zu	125	51,4	53,0	85,6
	trifft nicht zu	34	14,0	14,4	100,0
	Gesamt	236	97,1	100,0	
Fehlend	System	7	2,9		
Gesamt		243			

Messzeitpunkt 1
= vor der VO

spass_nachher NACH der LV - STATISTIK macht Spaß

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft zu	19	7,8	8,0	8,0
	trifft eher zu	99	40,7	41,6	49,6
	trifft eher nicht zu	84	34,6	35,3	84,9
	trifft nicht zu	36	14,8	15,1	100,0
	Gesamt	238	97,9	100,0	
Fehlend	System	5	2,1		
Gesamt		243			

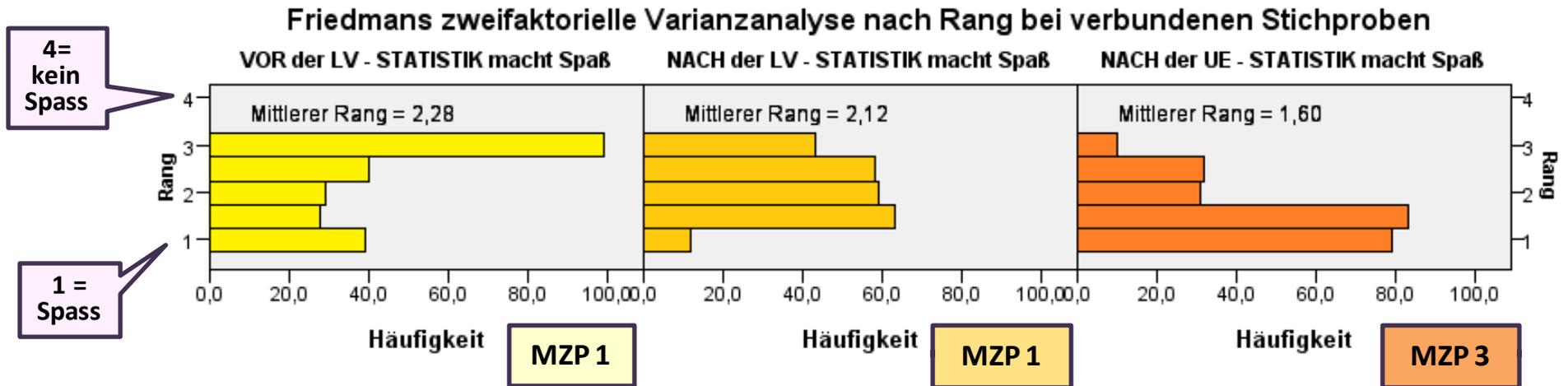
Messzeitpunkt 2
= nach der VO

spass_nach_UE NACH der UE - STATISTIK macht Spaß

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft zu	67	27,6	28,3	28,3
	trifft eher zu	106	43,6	44,7	73,0
	trifft eher nicht zu	49	20,2	20,7	93,7
	trifft nicht zu	15	6,2	6,3	100,0
	Gesamt	237	97,5	100,0	
Fehlend	System	6	2,5		
Gesamt		243	100,0		

Messzeitpunkt 3
= nach der UE

Friedman-Test



Gesamtanzahl	235
Teststatistik	78,763
Freiheitsgrade	2
Asymptotische Sig. (2-seitiger Test)	,000

Prüfmaß

WS vom Prüfmaß unter
Voraussetzung der Nullhypothese

Bei einer Fehler-WS von 0% wird die **Alternativhypothese (H_1)** angenommen:
Die Meinung, dass Statistik Spaß macht, nimmt zwischen Beginn der VO, Ende der VO und Ende der Übung signifikant zu ($F=78,763$, $n=235$, $p=0,000$).

Testen von Unterschiedshypothesen:

Parametrische Tests

t-Test für unabhängige Gruppen

einfaktorielle ANOVA

t-Test bei abhängigen Messwerten

VA für Messwiederholung

t-Test für unabhängige Stichproben: Vergleich **zweier**

Mittelwerte

→ Voraussetzungen:

- **metrisches** Datenniveau bei der Test-Variable (abhängige Variable)
- zwei Ausprägungen bei der Gruppenvariable (unabhängige Variable)
- Fallzahl in beiden Gruppen über 30
- oder: Test-Variable in beiden Gruppen annähernd **normalverteilt**

→ Nullhypothese:

Der durchschnittliche Testwert des Pluraltest unterscheidet sich nicht zwischen den Kindern aus Familien mit hohem bzw. niedrigem sozioökonomischen Status.



→ **Rechenlogik:** Vergleich der Mittelwerte beider Gruppen in Relation zur Streuung

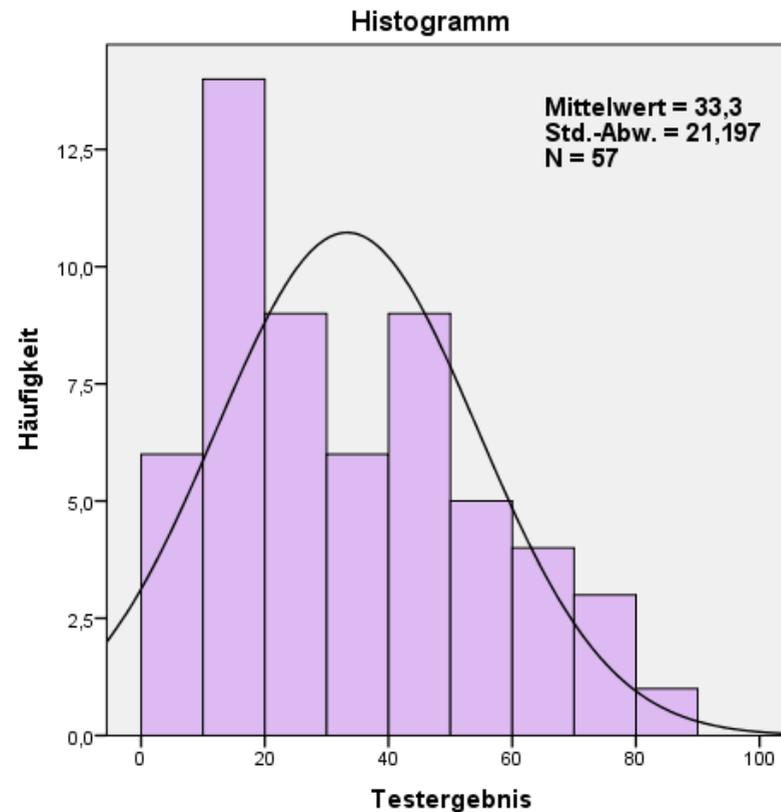
→ Prüfmaß = **t**

Beispiel: Metrische Testvariable und eine Gruppenvariable

Untersucht wird, ob sich der durchschnittliche Testwert des Pluraltests bei Kindern im Kindergartenalter nach dem sozioökonomischen Status der Herkunftsfamilie unterscheidet.

SES Sozioökonomischer Status (SES)

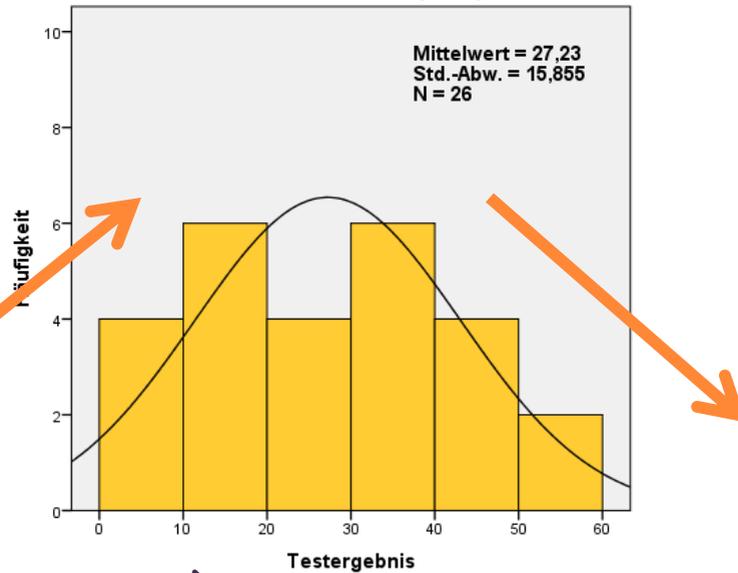
		Häufigkeit	Prozent
Gültig	Low SES	26	45,6
	High SES	31	54,4
	Gesamt	57	100,0



Prüfen auf Normalverteilung in den Subgruppen

Histogramm

Sozioökonomischer Status (SES): Low SES

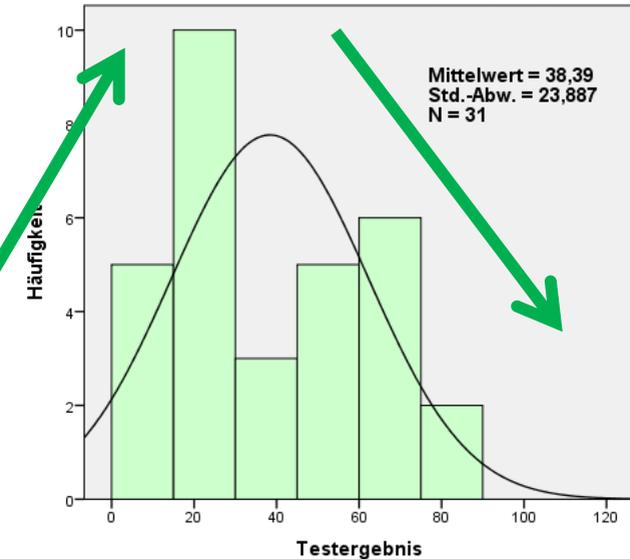


Gipfel annähernd
in der Mitte

Schiefe = 0,021
(zwischen 0 und ± 1)

Histogramm

Sozioökonomischer Status (SES): High SES



Gipfel annähernd
in der Mitte

Schiefe = 0,286
(zwischen 0 und ± 1)

t-Test

Gruppenstatistiken

	SES Sozioökonomischer Status (SES)	N	Mittelwert	Standardabweichung
ppvt1Dt_Rw Testergebnis	Low SES	26	27,23	15,855
	High SES	31	38,39	23,887

durchschnittliches Testergebnis bei den beiden Gruppen

Prüfmaß t = Vergleich der Mittelwerte

Test bei unabhängigen Stichproben

T-Test für die Mittelwertgleichheit

	T	df	Sig. (2-seitig)	Mittlere Differenz
ppvt1Dt_Rw Testergebnis	-2,034	55	,047	-11,156

WS von Prüfmaß t unter Voraussetzung der Nullhypothese

t-Test auf Mittelwertgleichheit

Bei einer Fehler-WS von 4,7% ($p \leq 5\%$) wird die **Alternativhypothese (H_A)** angenommen:

*Das durchschnittliche Testergebnis des Pluraltests unterscheidet sich **signifikant** nach dem sozioökonomischen Status der Herkunftsfamilie ($t=-2,034$, $df=55$, $p=0,047$).*

*Kinder aus Familien mit hohem SES erreichten ein **signifikant besseres durchschnittliches Testergebnis** (38,4 Pkt.) als jene mit niedrigem SES (27,2 Pkt.).*

Beispiel: Gilt dieser Unterschied gleichermaßen bei Familien mit Deutscher und Türkischer Erstsprache?

Gruppenstatistiken

Sprache Sprache der Eltern		SES Sozio- ökonomischer Status (SES)	N	Mittelwert	Standard- abweichung
Deutsch	ppvt1Dt_Rw Testergebnis	Low SES	14	36,79	11,995
		High SES	15	58,20	15,694
Türkisch	ppvt1Dt_Rw Testergebnis	Low SES	12	16,08	12,229
		High SES	16	19,81	12,276

Test bei unabhängigen Stichproben

T-Test für die Mittelwertgleichheit

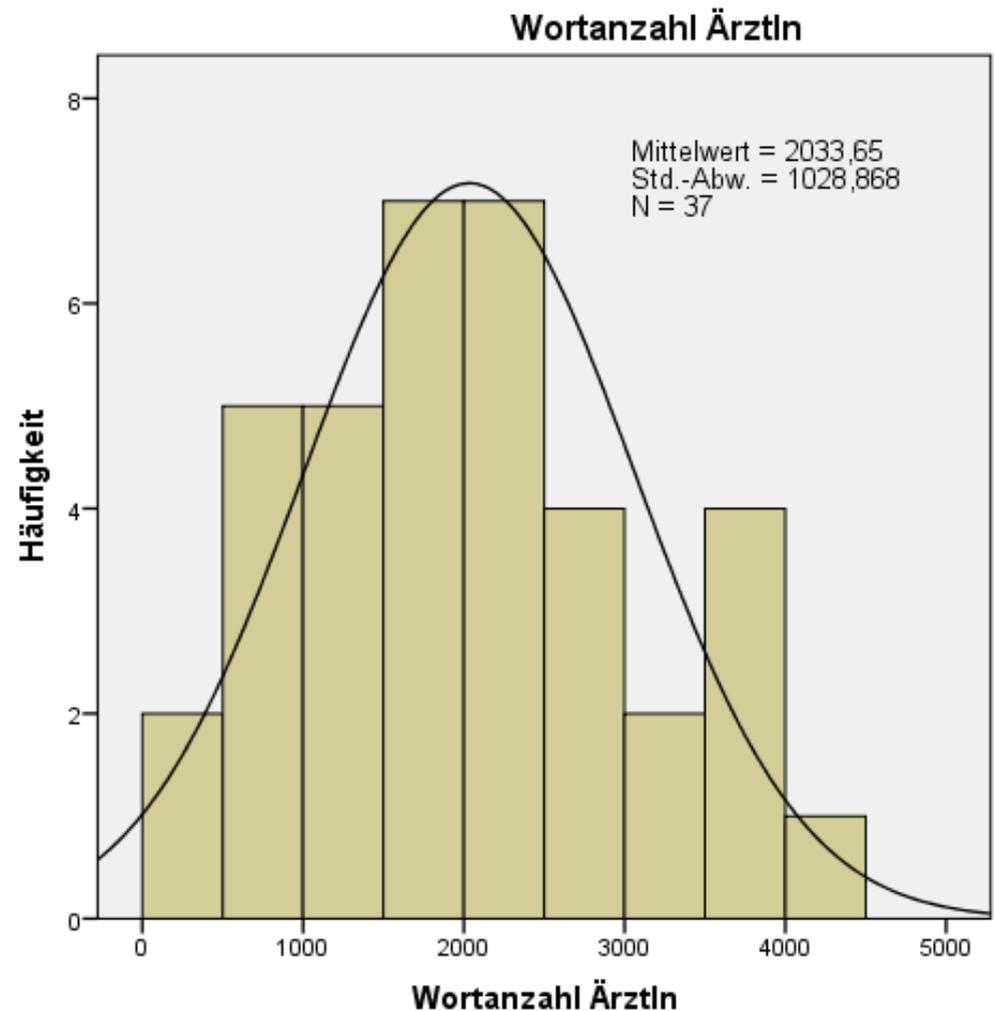
Sprache Sprache der Eltern		T	df	Sig. (2-seitig)	Mittlere Differenz
Deutsch	ppvt1Dt_Rw Testergebnis	-4,144	26,028	,000	-21,414
Türkisch	ppvt1Dt_Rw Testergebnis	-,797	26	,433	-3,729

Der durchschnittliche Testwert unterscheidet sich signifikant bei den Familien mit Deutscher Erstsprache nach SES.

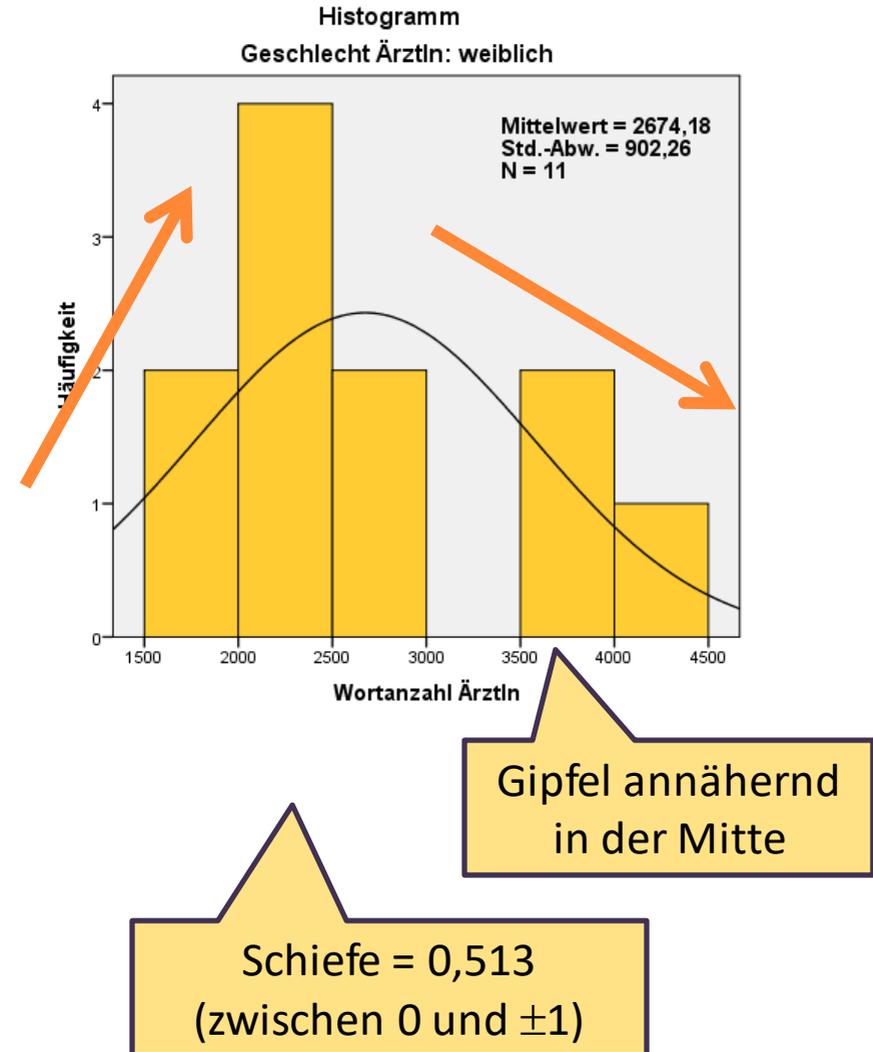
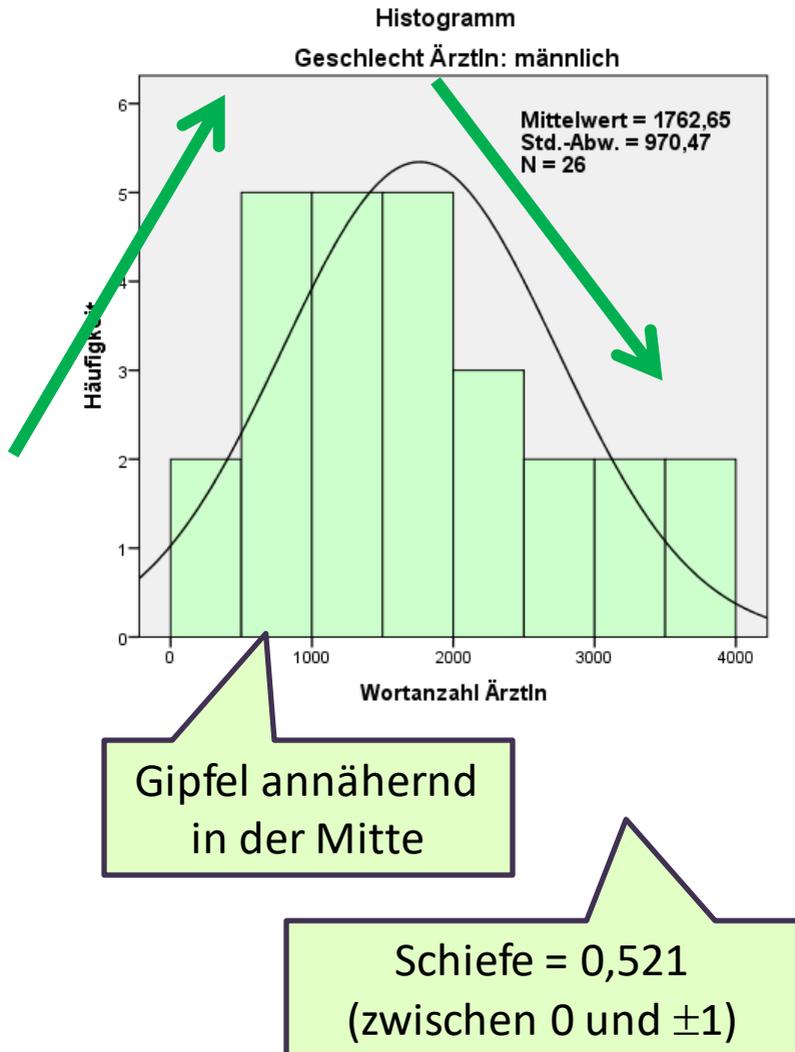
Bei den Familien mit Türkischer Erstsprache gibt es keinen signifikanten Unterschied nach SES.

Beispiel: Unterscheidet sich die durchschnittliche Wortanzahl zwischen Gesprächen mit Ärztinnen und Ärzten? (Kopfschmerzambulanz)

Wortanzahl_A		Wortanzahl ÄrztIn			
	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente	
Gültig	480	1	1,3	2,7	2,7
	493	1	1,3	2,7	5,4
	542	1	1,3	2,7	8,1
	675	1	1,3	2,7	10,8
	748	1	1,3	2,7	13,5
	752	1	1,3	2,7	16,2
	897	1	1,3	2,7	18,9
	1101	1	1,3	2,7	21,6
	1256	1	1,3	2,7	24,3
	1310	1	1,3	2,7	27,0
	1446	1	1,3	2,7	29,7
	1461	1	1,3	2,7	32,4
	1502	1	1,3	2,7	35,1
	1674	1	1,3	2,7	37,8
	1676	1	1,3	2,7	40,5
	1752	1	1,3	2,7	43,2
	1762	1	1,3	2,7	45,9
	1886	1	1,3	2,7	48,6
	1930	1	1,3	2,7	51,4
	2028	1	1,3	2,7	54,1
	2064	1	1,3	2,7	56,8
	2145	2	2,7	5,4	62,2
	2257	1	1,3	2,7	64,9
	2276	1	1,3	2,7	67,6
	2459	1	1,3	2,7	70,3
	2579	1	1,3	2,7	73,0
	2727	1	1,3	2,7	75,7
	2908	1	1,3	2,7	78,4
	2914	1	1,3	2,7	81,1
	3001	1	1,3	2,7	83,8
	3434	1	1,3	2,7	86,5
	3506	1	1,3	2,7	89,2
	3610	1	1,3	2,7	91,9
	3744	1	1,3	2,7	94,6
	3996	1	1,3	2,7	97,3
	4109	1	1,3	2,7	100,0
	Gesamt	37	49,3	100,0	
Fehlend	99 keine Angabe	38	50,7		
Gesamt		75	100,0		



Beispiel: Unterscheidet sich die durchschnittliche Wortanzahl zwischen Gesprächen mit Ärztinnen und Ärzten? (Kopfschmerzambulanz)



t-Test

Gruppenstatistiken

	sex_A Geschlecht ÄrztIn	N	Mittelwert	Standard- abweichung
Wortanzahl ÄrztIn	männlich	26	1762,65	970,470
	weiblich	11	2674,18	902,260

Prüfmaß t

Test bei unabhängigen Stichproben

T-Test für die Mittelwertgleichheit

	T	df	Sig. (2-seitig)	Mittlere Differenz
Wortanzahl_A Wortanzahl ÄrztIn	-2,663	35	,012	-911,528

Fehler-WS = 1,2%

Bei einer Fehler-WS von 1,2% wird die **Alternativhypothese (H_A) angenommen**:
In der Kopfschmerzambulanz werden bei Gesprächen mit männlichen Ärzten durchschnittlich signifikant weniger Wörter vom Arzt gesprochen (1763 Wörter) als bei Gesprächen mit Ärztinnen (2674 Wörter) (t=-2,663, df=35, p=0,012).

Einfaktorielle Varianzanalyse: Vergleich **mehrerer** Mittelwerte

→ Voraussetzungen:

- **metrisches** Datenniveau bei der Test-Variable (abhängige Variable)
- kategoriale Gruppenvariable (unabhängige Variable)
- Fallzahl in den Gruppen über 30
- oder: Test-Variable in den Gruppen annähernd **normalverteilt**

→ Nullhypothese:

Der durchschnittliche Testwert des Pluraltest unterscheidet sich nicht zwischen den untersuchten Gruppen.



→ **Rechenlogik:** Die Streuung der Gruppenmittelwerte um den Gesamtmittelwert ist größer als die individuelle Streuung innerhalb der Gruppen (= multipler t-Test)

→ Prüfmaß = **F**

ANOVA

ONEWAY deskriptive Statistiken

ppvt1Dt_Rw Testergebnis

	N	Mittelwert	Standard- abweichung	95%-Konfidenzintervall für den Mittelwert	
				Untergrenze	Obergrenze
High SES - Deutsch	15	58,20	15,694	49,51	66,89
Low SES - Deutsch	14	36,79	11,995	29,86	43,71
High SES - Türkisch	16	19,81	12,276	13,27	26,35
Low SES - Türkisch	12	16,08	12,229	8,31	23,85
Gesamt	57	33,30	21,197	27,67	38,92

Es besteht ein signifikanter Unterschied im durchschnittlichen Testergebnis zwischen den Subgruppen.

Einfaktorielle ANOVA

ppvt1Dt_Rw Testergebnis

	Quadrat- summe	df	Mittel der Quadrat e	F	Signifikanz
Zwischen den Gruppen	15937,819	3	5312,61	30,525	,000
Innerhalb der Gruppen	9224,111	53	174,040		
Gesamt	25161,930	56			

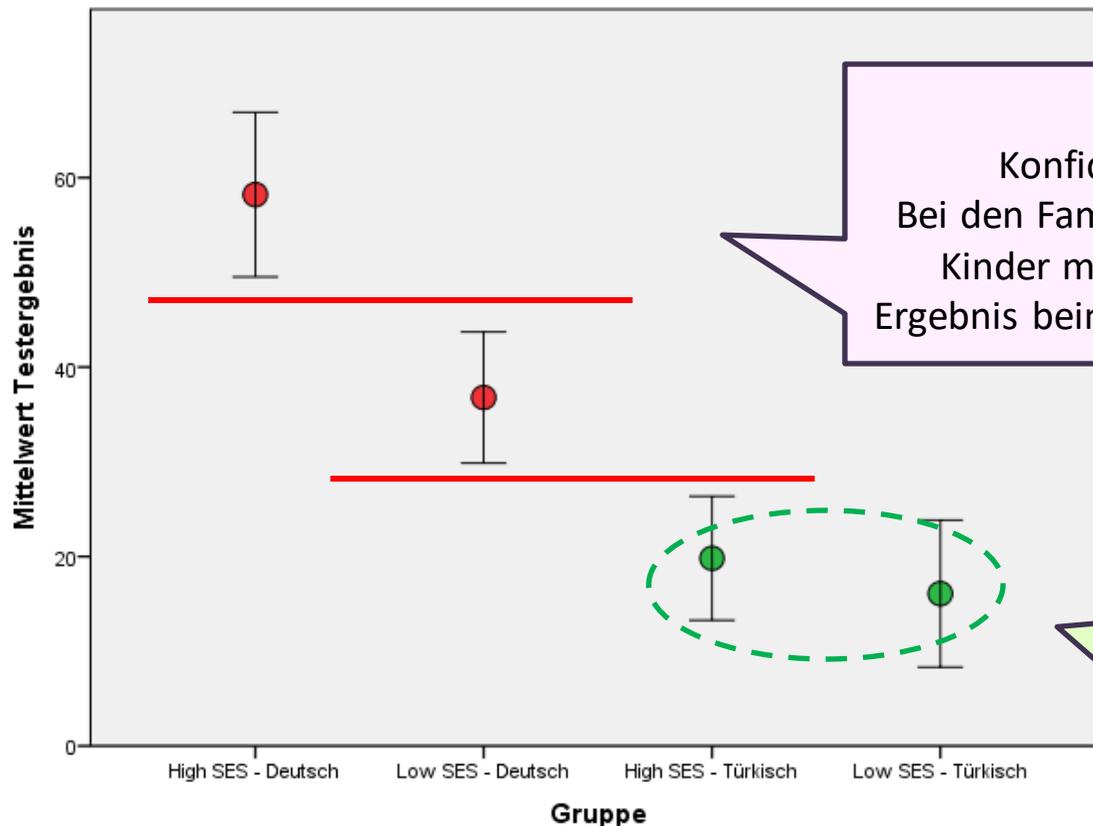
Prüfmaß F

Wahrscheinlichkeit des Prüfmaß F bei Zufall

Bei einer Fehler-WS von 0% wird die **Alternativhypothese (H_A)** angenommen:

Zwischen den vier Vergleichsgruppen besteht ein signifikanter Unterschied hinsichtlich des durchschnittlichen Testergebnis beim Pluraltest:

Kinder aus Familien mit deutscher Erstsprache erreichen signifikant bessere Ergebnisse im Pluraltest als Kinder mit türkischer Erstsprache ($F=30,525$, $n=57$, $p=0,000$).



Erstsprache Deutsch:

Konfidenzintervalle überschneiden sich nicht. Bei den Familien mit deutscher Erstsprache erreichen die Kinder mit dem höheren SES ein signifikant besseres Ergebnis beim Pluraltest als die Kinder mit niedrigerem SES.

Erstsprache Türkisch:

Konfidenzintervalle überschneiden sich. Bei den Familien mit türkischer Erstsprache zeigen sich keine signifikanten Unterschiede im Pluraltestergebnis nach SES.

t-Test für abhängige Stichproben: Vergleich der Mittelwerte einer Veränderung/Pairung

→ Voraussetzungen:

- zwei Testvariablen, die inhaltlich voneinander abhängig sind
- **metrisches** Datenniveau der Test-Variablen
- Fallzahl über 30
- oder: Differenz der Test-Variablen annähernd **normalverteilt**

→ Nullhypothese:

Das durchschnittliche Testergebnis des Wortschatztests verändert sich nicht zwischen Beginn und Ende der Maßnahme.

→ **Rechenlogik:** Berechnen der Differenzen zwischen den Messwerten zu beiden Messzeitpunkten:
Weicht der Mittelwert der Differenz signifikant von Null ab?

→ Prüfmaß = **t**

Beispiel: Maßnahme in Kindergärten zur Sprachförderung

Finanzierung für eine Kindergartengruppe von 30 Kindern

Begleitende Evaluation: Wortschatztest zu zwei Zeitpunkten

Fragestellung: **Wirkt die Fördermaßnahme?**

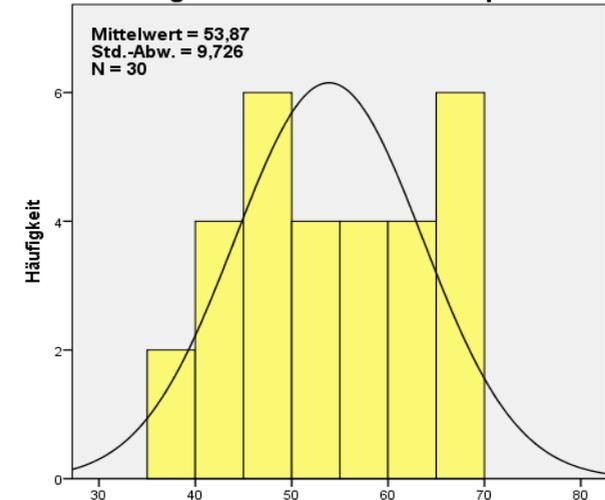
Statistik bei gepaarten Stichproben

	Mittelwert	N	Standard- abweichung
mzp1 Testergebnis MP1	53,87	30	9,726
mzp2 Testergebnis MP2	54,20	30	7,748

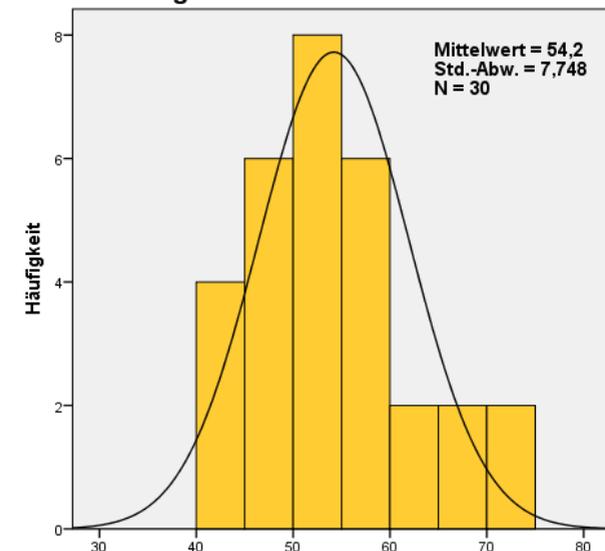
Durchschnittlicher
Testwert bei der
ersten Messung
Beginn der
Maßnahme =
September

Durchschnittlicher
Testwert bei der
zweiten Messung
Ende der
Maßnahme =
Dezember

Testergebnis Wortschatztest September



Testergebnis Wortschatztest Dezember



t-Test bei gepaarten Stichproben

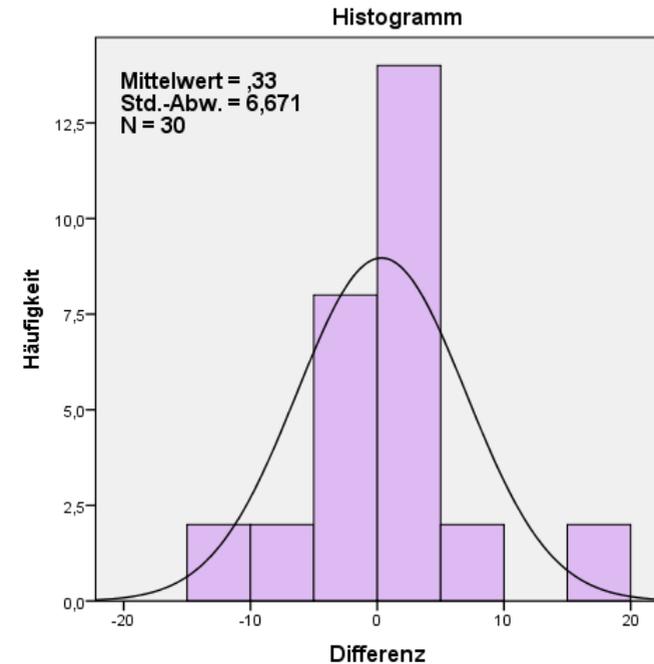
Bei zwei Kindern hat sich der Testwert um 13 Punkte verschlechtert.

Bei zwei Kindern ist der Testwert gleich geblieben.

Bei zwei Kindern hat sich der Testwert um 15 Punkte verbessert.

	Differenz			
	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
	-13	2	6,3	6,7
	-10	2	6,3	13,3
	-4	4	12,5	26,7
	-3	2	6,3	33,3
	-2	2	6,3	40,0
	0	2	6,3	46,7
	2	4	12,5	60,0
	4	8	25,0	86,7
	6	2	6,3	93,3
	15	2	6,3	100,0
Gesamt		30	93,8	100,0
Fehlend	System	2	6,3	
Gesamt		32	100,0	

Die durchschnittliche Veränderung des Testergebnis beträgt + 0,33 Punkte.



t-Test bei gepaarten Stichproben

Korrelationen bei gepaarten Stichproben

	N	Korrelation	Signifikanz
mzp1 Testergebnis MP1 & mzp2 Tesergebnis MP2	30	,731	,000

Hohe positive Korrelation – inhaltliche Abhängigkeit ist gegeben.

Test bei gepaarten Stichproben

	Gepaarte Differenzen		T	df	Sig. (2-seitig)
	Mittelwert	Standardabweichung			
mzp1 Testergebnis MP1 - mzp2 Tesergebnis MP2	,333	6,671	,274	29	,786

Prüfmaß t

Die Fehler-WS beträgt 78,6%.

Die durchschnittliche Veränderung des Testergebnis beträgt + 0,33 Punkte.

Bei einer Fehlerwahrscheinlichkeit von $> 5\%$ wird die **Nullhypothese (H_0) beibehalten**:
Es zeigt sich keine signifikante Veränderung (bzw. Verbesserung) des durchschnittlichen Testergebnisses beim Wortschatztest der Kinder zwischen Beginn und Ende der Maßnahme ($t=0,274$, $df=29$, $p=786$).

Varianzanalyse bei Messwiederholung: Vergleich der Mittelwerte **mehrerer** Messzeitpunkte

→ Voraussetzungen:

- mehrere Testvariablen, die inhaltlich voneinander abhängig sind
- **metrisches** Datenniveau der Test-Variablen
- Fallzahl über 30
- oder: Differenzen der Test-Variablen annähernd **normalverteilt**

→ Nullhypothese:

Das durchschnittliche Testergebnis des Wortschatztests verändert sich nicht zwischen den Messzeitpunkten.

→ **Rechenlogik:** Die Streuung der durchschnittlichen Testergebnisse zu den Messzeitpunkten ist größer als die individuelle Streuung (= multipler t-Test)

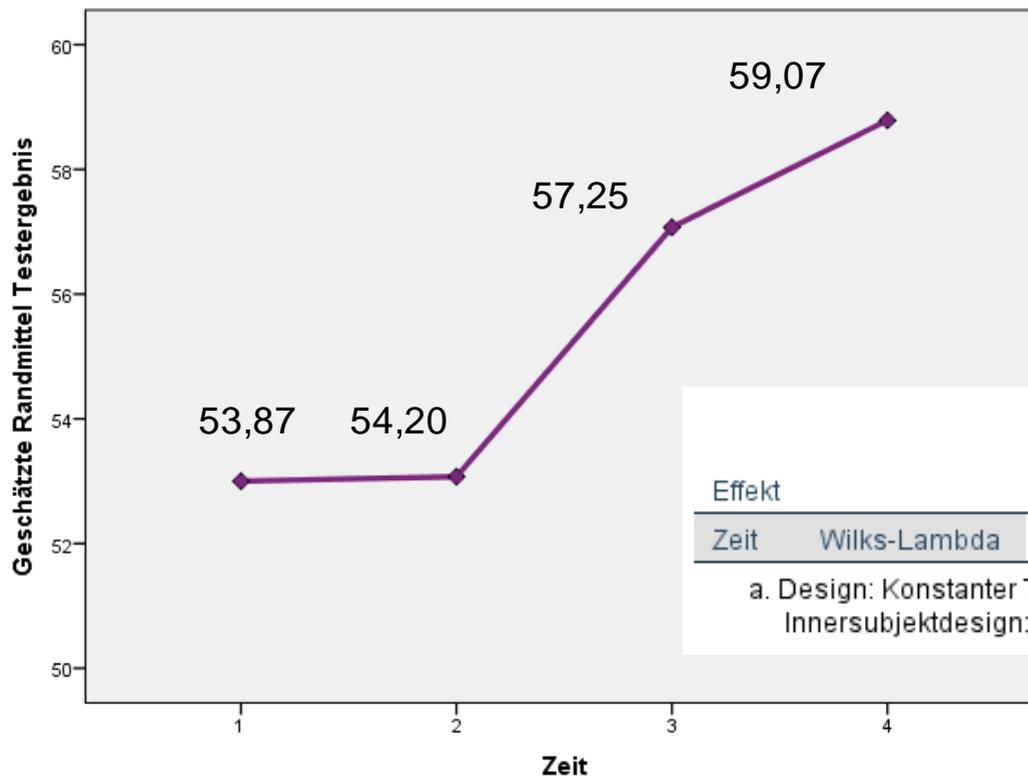
→ Prüfmaß = **F** (Wilks Lambda)

Beispiel: Maßnahme in Kindergärten zur Sprachförderung

Die Maßnahme wird verlängert → Wortschatztest zu vier Zeitpunkten

Fragestellung: Wirkt die Fördermaßnahme?

Varianzanalyse bei Messwiederholung



Signifikante Veränderung zwischen den vier Messzeitpunkten.

Multivariate Tests^a

Effekt	Wert	F	Hypothese df	Fehler df	Sig.
Zeit Wilks-Lambda	,572	6,242	3,000	25,000	,003

a. Design: Konstanter Term
Innersubjektdesign: Zeit

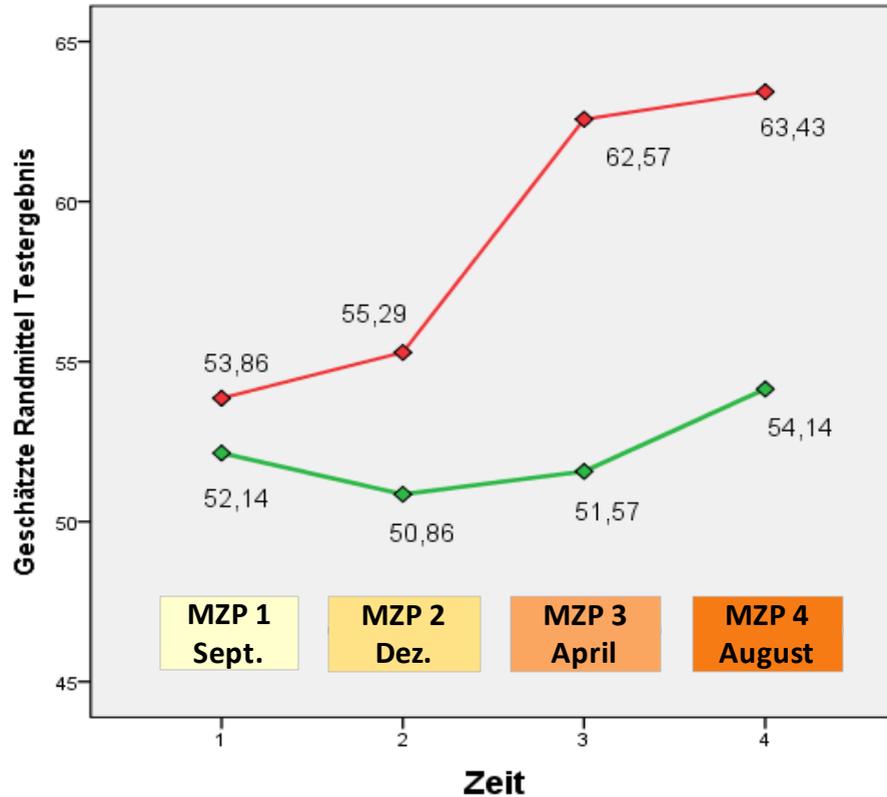
MZP 1
Sept.

MZP 2
Dez.

MZP 3
April

MZP 4
August

Varianzanalyse bei Messwiederholung – Vergleich nach SES



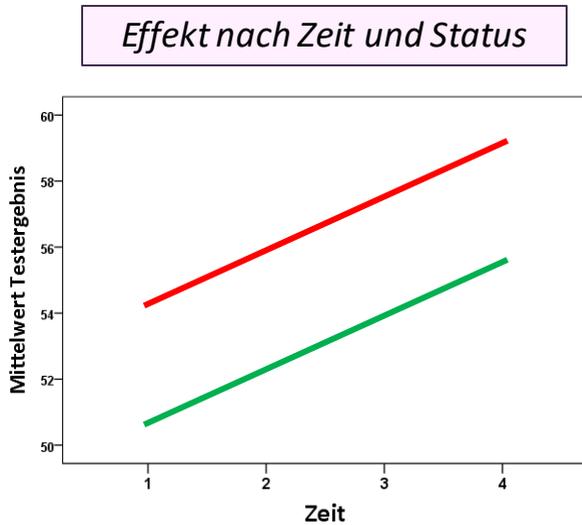
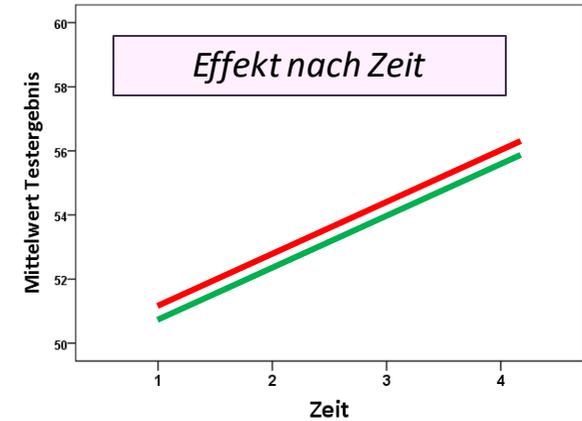
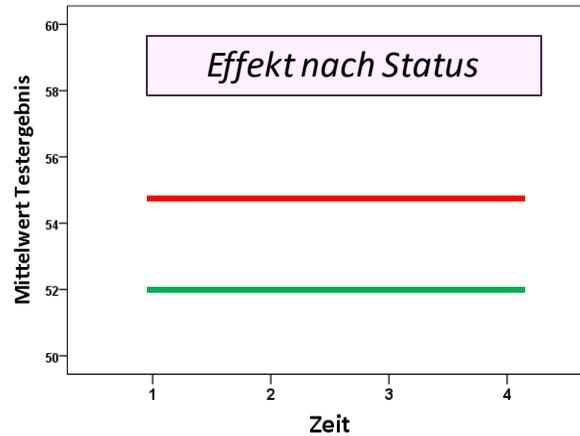
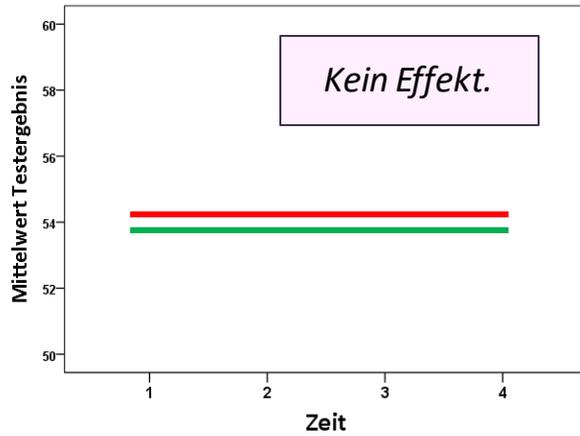
Veränderungen nach Zeit insgesamt signifikant

Interaktion zwischen Faktor Zeit und Faktor Gruppe ist signifikant.

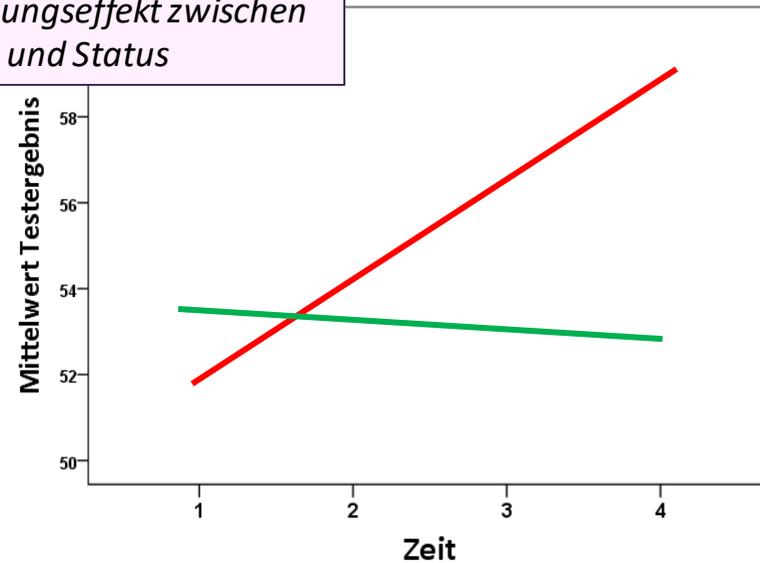
Multivariate Tests^a

Effekt	Wert	F	Hypothese df	Fehler df	Sig.	
zeit	Wilks-Lambda	,513	7,580	3,000	24,000	,001
zeit * SES	Wilks-Lambda	,660	4,115	3,000	24,000	,017

a. Design: Konstanter Term + SES
Innersubjektdesign: zeit



Interaktion von Zeit mit Status = Wechselwirkungseffekt zwischen Zeit und Status



SES
— High
— Low

Übersicht über Signifikanztests

Daten	Test	Voraussetzungen	Ergebnis	Bsp. zur Nullhypothese
Kreuz-tabelle (ab nominal)	Chi-Quadrat	In jeder Zelle (bzw. in 80% der Zellen) beträgt die erwartete Hfkt. mind. 5.	Sign. Unterschied der tatsächlichen und erwarteten Häufigkeiten	Es besteht kein Unterschied nach Geschlecht und Anzahl der Wörter in Kategorien.
ordinal und zwei Gruppen	U-Test	Testvariable: - zumindest ordinal - metrisch, aber nicht normalverteilt	Sign. Unterschied der Verteilung der Testvariable in beiden Gruppen	Die Wichtigkeit von Grammatik bei Simultanübersetzungen unterscheidet sich nicht zwischen zwei Berufsgruppen.
ordinal und abhängig	Wilcoxon- Test	Testvariablen: - zumindest ordinal - metrisch, aber nicht normalverteilt	Sign. Unterschied der Verteilung vorher - nachher	Die Einschätzung, dass Statistik Spass macht unterscheidet sich nicht zwischen Beginn und Ende der LV.
metrisch und zwei Gruppen	t-Test für unabh. Stpr.	Testvariable: metrisch und normalverteilt - bzw. Fallzahl pro Gruppe > 30	Sign. Unterschied der Mittelwerte der Testvariable in beiden Gruppen	Die durchschnittliche Wortanzahl unterscheidet sich nicht nach Geschlecht.
metrisch und abhängig	t-Test für abh. Stpr.	Testvariablen: metrisch und normalverteilt - bzw. Fallzahl > 30	Sign. Unterschied der Mittelwerte vorher - nachher	Das durchschnittliche Testergebnis unterscheidet sich nicht zwischen Beginn und Ende der Maßnahme.

Schritt 1:
Welches Datenniveau hat die abhängige Variable?

Schritt 2:
Was soll untersucht werden?

Schritt 3:
Testauswahl



Korrelation
= Stärke des Zusammenhangs

Korrelation = Zusammenhang zwischen zwei Variablen (bivariat)

= Maß für Stärke und Richtung eines Zusammenhangs zweier Variablen

Korrelationskoeffizient:

Das Ausmaß der Stärke ist normiert zwischen -1 0 +1.

Dadurch kann zwischen stärkeren und schwächeren Zusammenhängen differiert werden.

Korrelation ermöglicht also den Vergleich von Abhängigkeiten:

„Die **Arbeitszufriedenheit** hängt **besonders stark** von der Zusammenarbeit im Team ab.

Sie hängt auch, aber **weniger stark**, vom Verhältnis zur/m Vorgesetzten und der Belastung durch Monotonie ab.

Ungünstige Arbeitshaltung oder die Platz- und Lichtverhältnisse beeinflussen die Arbeitszufriedenheit **kaum**.“

Korrelation – Stärke und Richtung eines Zusammenhanges

→ **positiver Zusammenhang (maximal +1):**

Je **höher** die Ausprägung in Variable A, desto **höher** die Ausprägung in Variable B.

Bsp: *"Je älter die Befragten, desto höher ist deren Einkommen."*

→ **kein Zusammenhang (nahe bei 0):**

Die Ausprägungen von Variable A sind **unabhängig** von jenen von Variable B.

Bsp: *"Wie oft die Befragten ins Kino gehen hat nichts mit deren Einkommen zu tun."*
Oder: *„Die Häufigkeit ins Kino zu gehen ist unabhängig vom Einkommen“.*

→ **negativer Zusammenhang (maximal -1):**

Je **höher** die Ausprägung in Variable A, desto **niedriger** die Ausprägung in Variable B.

Bsp: *"Je älter die Befragten, desto seltener gehen sie ins Kino."*

Korrelation = Zusammenhang zwischen zwei Variablen (bivariat)

„Das Strickmuster erkennen“

	Note	
	Mathe	Physik
Susi	1	1
Trixi	2	2
Max	3	3
Eva	4	4
Otto	5	5

**Positiver
Zusammenhang**
"Glatt"

Je besser die Note in
Mathe, desto besser
die Note in Physik.

**Maximale
Korrelation**

+1

	Note	
	Mathe	Deutsch
Susi	1	5
Trixi	2	4
Max	3	3
Eva	4	2
Otto	5	1

**Negativer
Zusammenhang**
"Verkehrt"

Je besser die Note in
Mathe, desto schlechter
die Note in Deutsch.

**Maximale
Korrelation**

-1

	Note	
	Mathe	Turnen
Susi	1	1
Trixi	2	2
Max	3	3
Eva	4	2
Otto	5	1

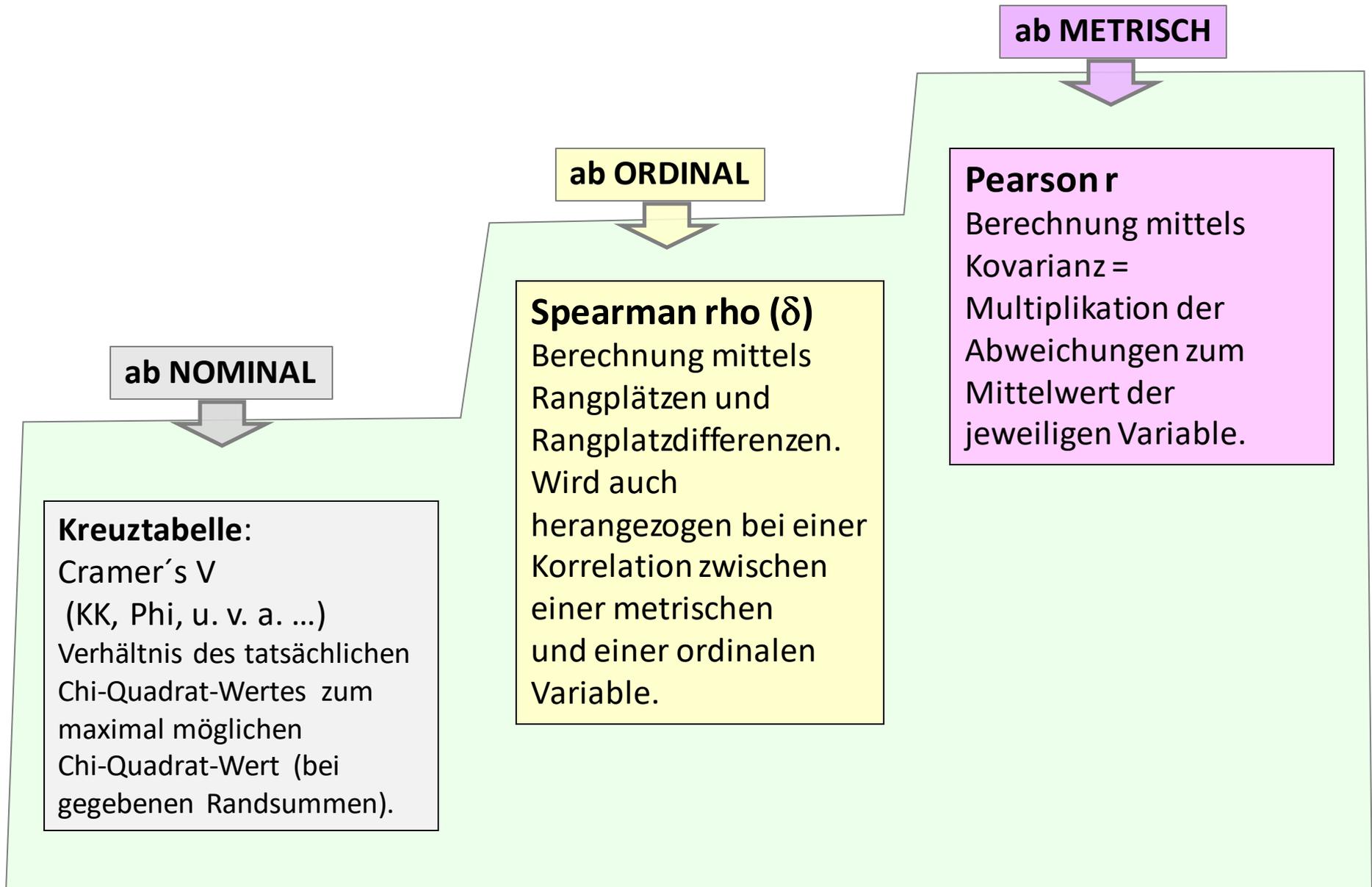
**Kein
Zusammenhang**
"Glatt und verkehrt"

Die Note in Mathe hat
nichts zu tun mit der
Note in Turnen.

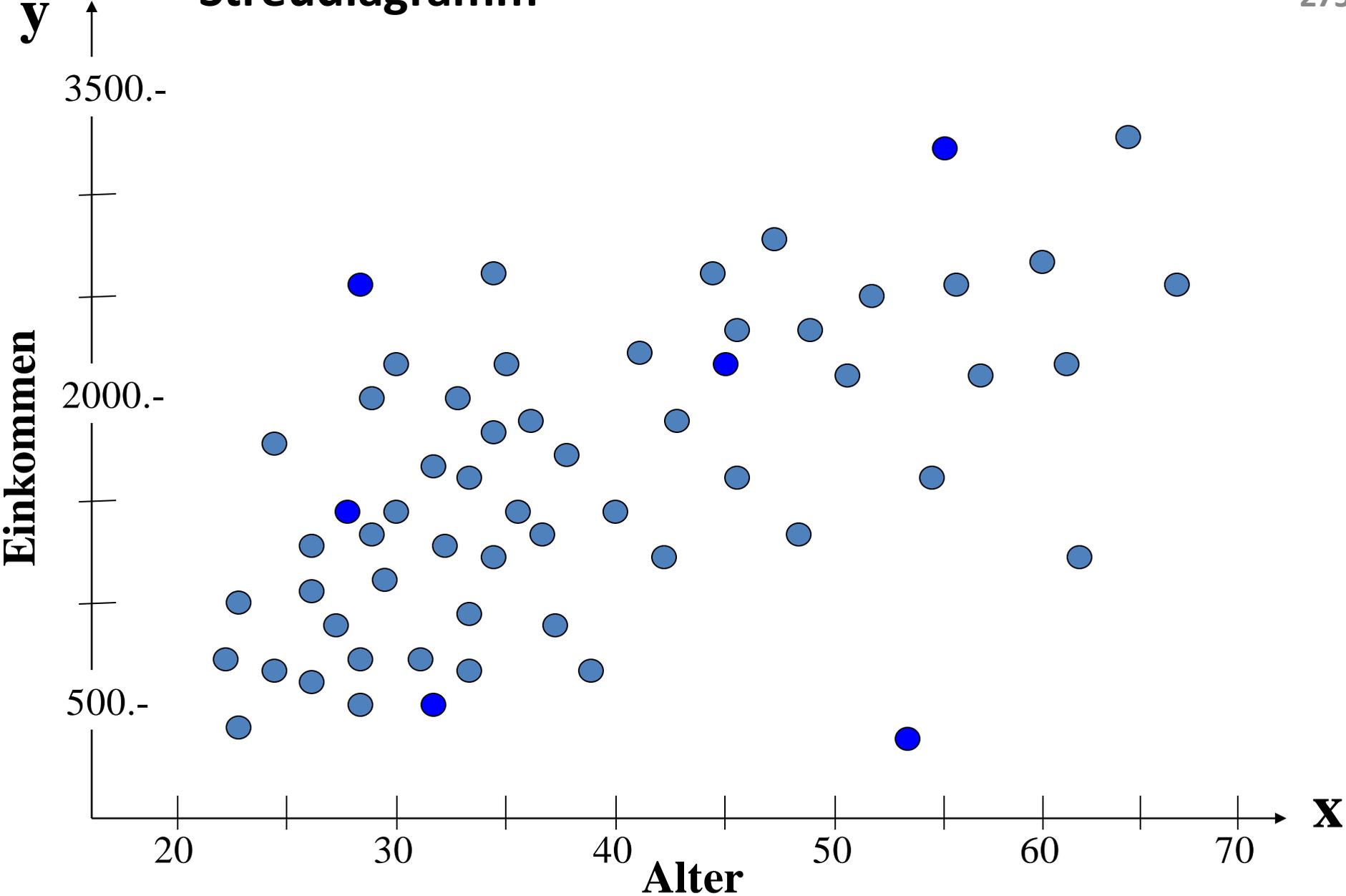
**Minimale
Korrelation**

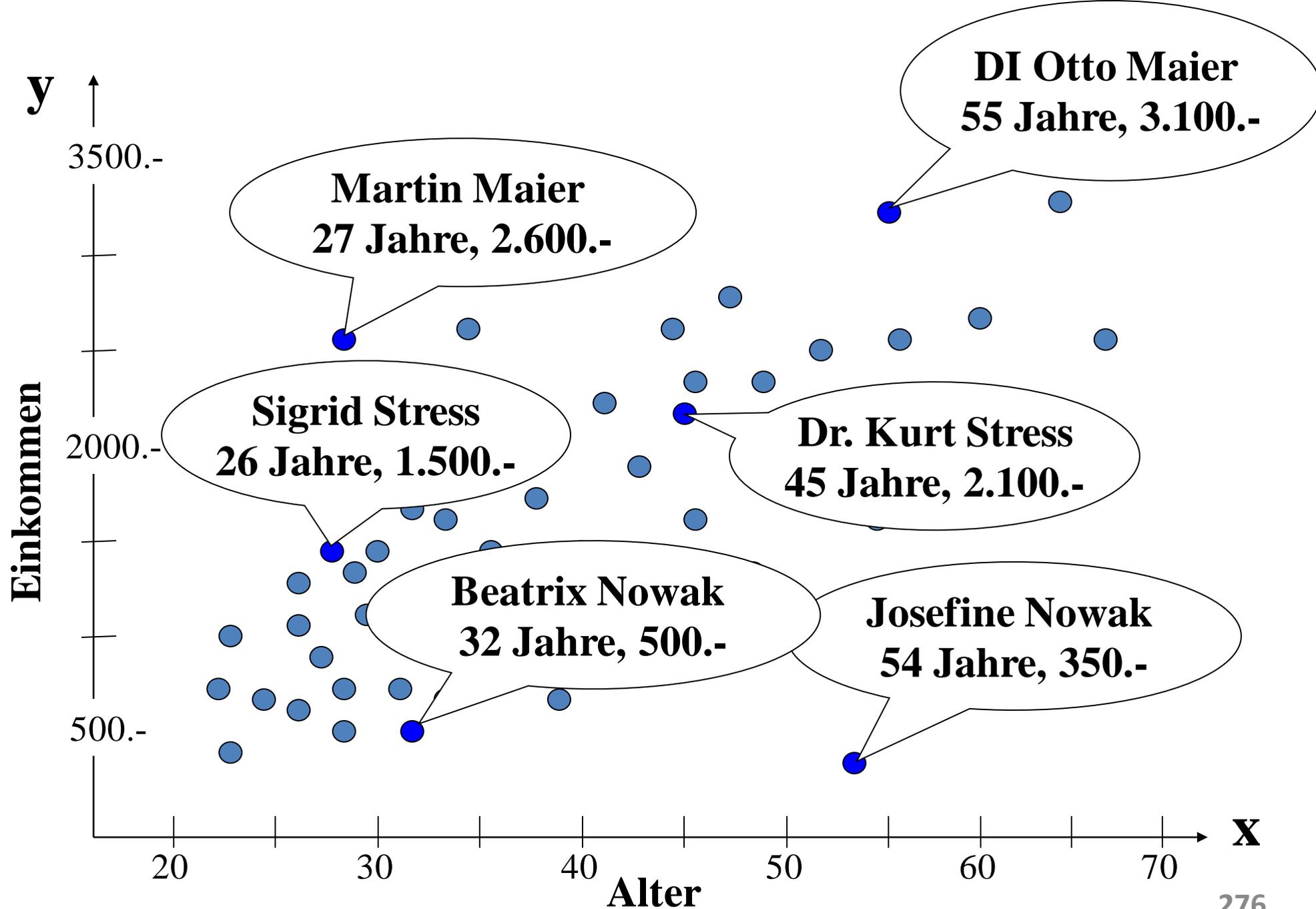
0

Korrelation und Datenniveau

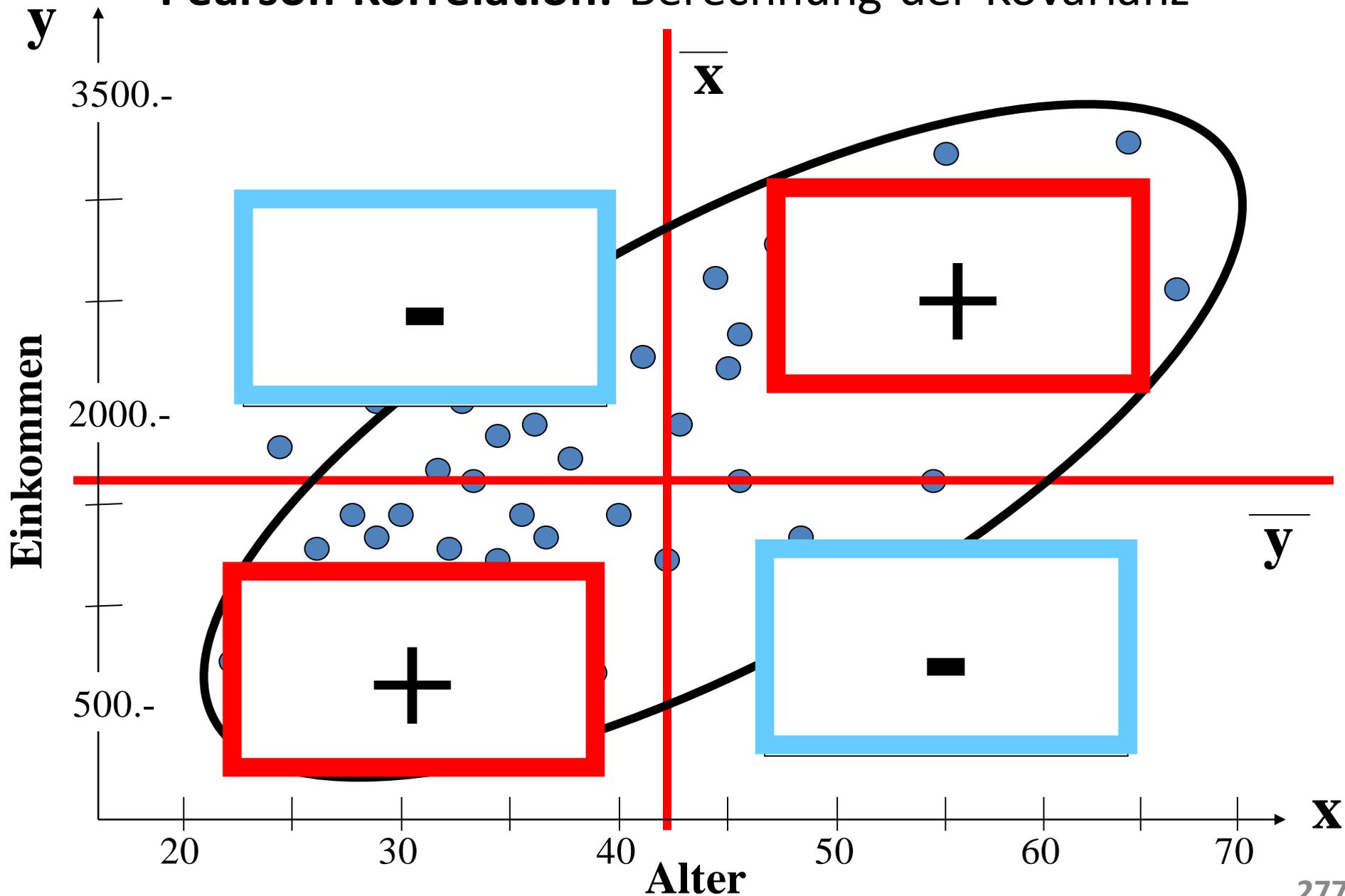


Streudiagramm





Pearson-Korrelation: Berechnung der Kovarianz



Kovarianz xy

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$r = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Standardabweichung von Variable x

Standardabweichung von Variable y

Pearson-Korrelation: Berechnung der Korrelation

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{S_{xy}}{S_x S_y}$$

$$\phi = \frac{a \cdot d - b \cdot c}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}$$

	A=0	A=1	Summe
B=0	a	b	a+b
B=1	c	d	c+d
Summe	a+c	b+d	a+b+c+d

Zugrunde liegendes Modell: es besteht ein **linearer** Zusammenhang zwischen zwei Merkmalen X und Y, in einem Streudiagramm der beiden Merkmale lässt sich eine Gerade in die Punktwolke einzeichnen.

Die Korrelation nach Pearson misst, wie **eng** die Punkte um die Gerade liegen. (Und nicht wie steil die Gerade verläuft)

Pearsonkorrelation

Um den Zusammenhang zwischen zwei metrisch skalierten Merkmalen zu messen, wird der Pearsonkorrelationskoeffizient verwendet.

Die Berechnung beruht auf der Berechnung der Kovarianz und der anschließenden Normierung derselben.

Die Kovarianz kann sehr hohe Werte annehmen und ist nicht interpretierbar. Durch die Normierung (Division durch $s_x s_y$) liegt der Wertebereich dann zwischen -1 und +1.

Anmerkung: Kovarianz = r, wenn:

... man die Kovarianz aus zwei standardisierten Variablen berechnet, dann liegt diese ebenfalls zwischen -1 und +1.

Standardisierung einer Variable: $X \rightarrow Z_X: z_i = \frac{x_i - \bar{x}}{s}$

Pearson-Korrelation: Berechnung in Excel

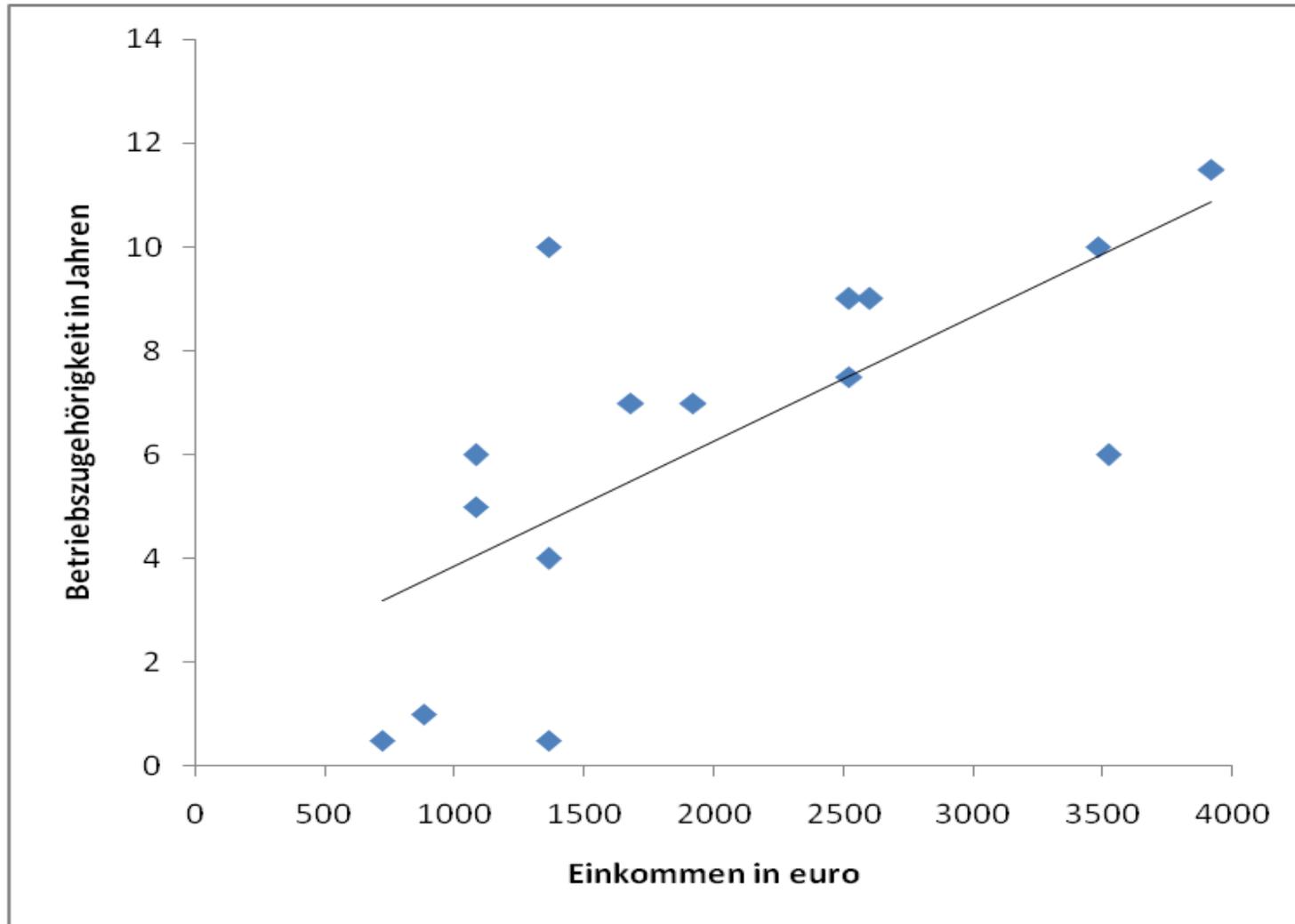
	A	B	C	D	E	F	G	H	I	J	K
1	Pearson-Korrelation Einkommen mit Dauer Betriebszugehörigkeit										
2	nr	X euro	Y Betr.z	①	②	(x-xq)* (y-yq)	(x-xq) ²	(y-yq) ²			
3	1	1680	7	-320	0,73	-234,667	102400	0,53778			
4	2	2520	7,5	520	1,23	641,333	270400	1,52111	Kov. =	2401,33	
5	3	1360	4	-640	-2,27	1450,67	409600	5,13778	Koeff r =	0,70493	
6	4	2600	9	600	2,73	1640	360000	7,47111		③	
7	5	3480	10	1480	3,73	5525,33	2190400	13,9378			
8	6	1920	7	-80	0,73	-58,6667	6400	0,53778	SPSS r =	0,70493	
9	7	1360	10	-640	3,73	-2389,33	409600	13,9378			
10	8	720	0,5	-1280	-5,77	7381,33	1638400	33,2544			
11	9	3920	11,5	1920	5,23	10048	3686400	27,3878			
12	10	1080	5	-920	-1,27	1165,33	846400	1,60444			
13	11	1360	0,5	-640	-5,77	3690,67	409600	33,2544			
14	12	2520	9	520	2,73	1421,33	270400	7,47111			
15	13	1080	6	-920	-0,27	245,333	846400	0,07111			
16	14	3520	6	1520	-0,27	-405,333	2310400	0,07111			
17	15	880	1	-1120	-5,27	5898,67	1254400	27,7378			
18	summ	30000	94	0	0	36020	15011200	173,933			
19	summ/n	2000	6,2667			2401,33	1000747	11,5956			
20	s (Stdabw.)						1000,373	3,40522			

③ = F19 / (G20 * H20)

Wichtigste Formeln in Excel (Mit dem \$-Zeichen werden Spalten/Zeilen festgesetzt):
 Spalte D (für ersten Fall) ① = B3-\$B\$19 Spalte E (für ersten Fall) ② = C3-\$C\$19

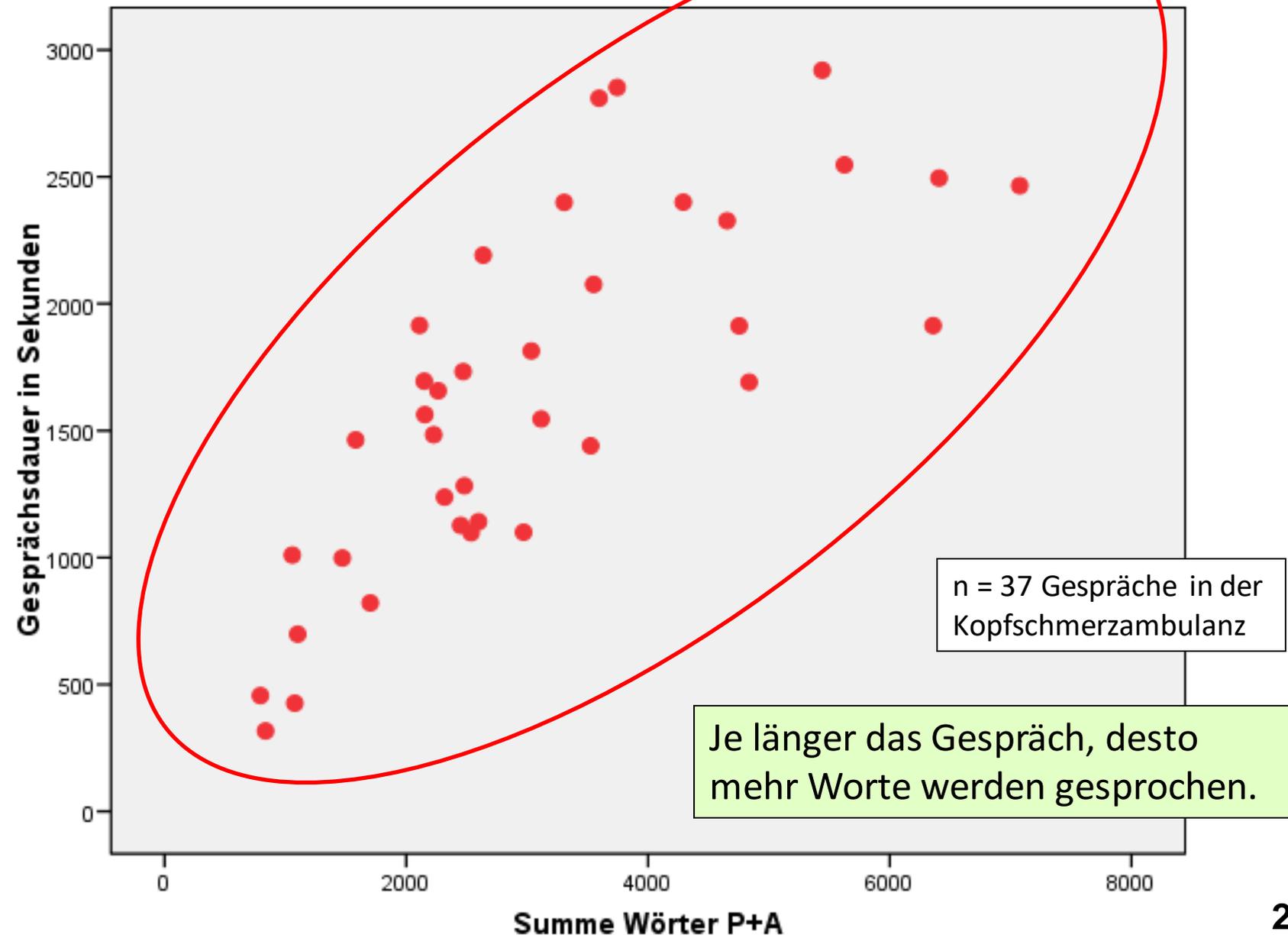
Pearson-Korrelation:

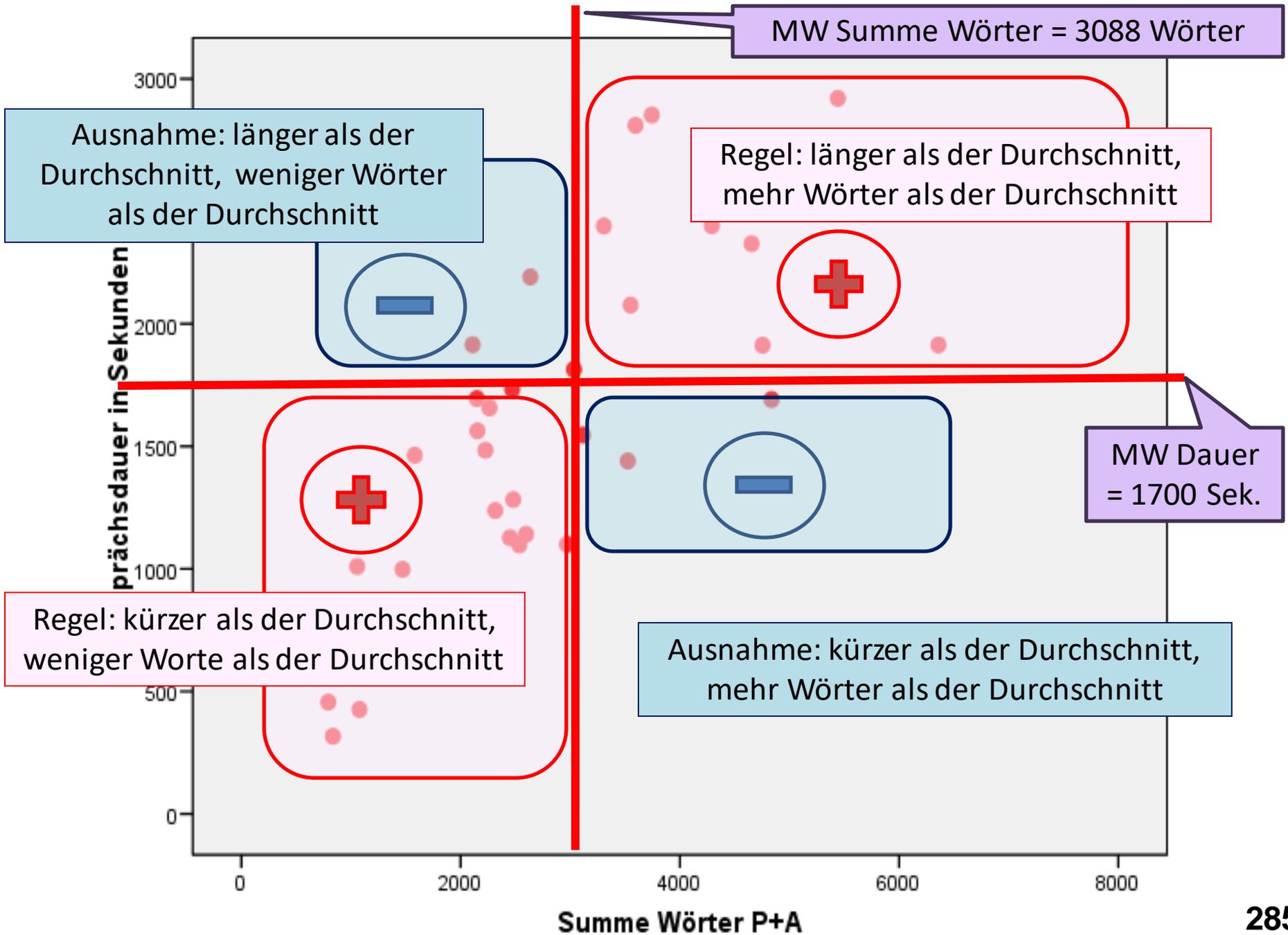
Streudiagramm zu obigem Rechenbeispiel

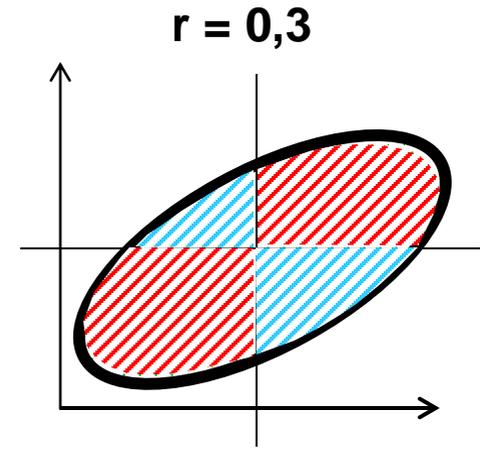
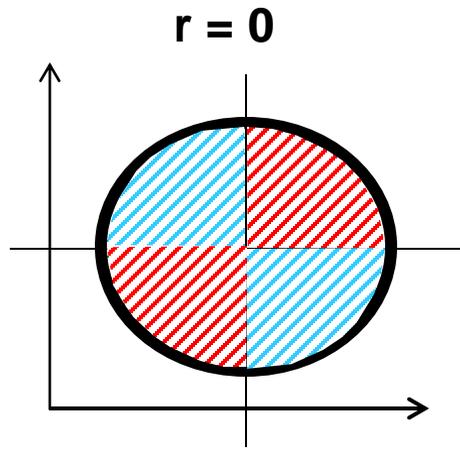
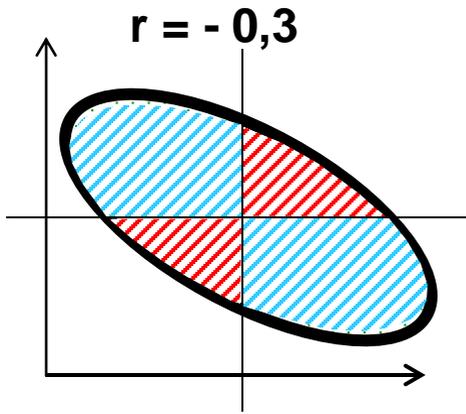


Checkliste zur Berechnung von Korrelationen

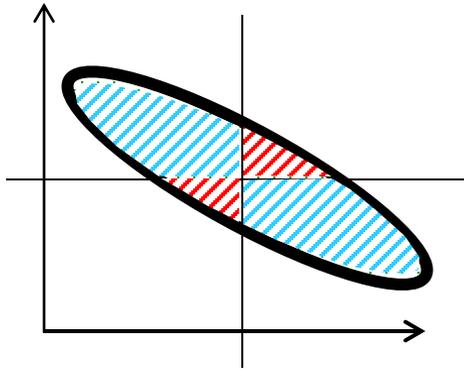
1. Vorher: **missing values** als solche definieren! (99 = k.a.)
2. Linearität des Zusammenhangs überprüfen:
 - a) in der Theorie (macht das überhaupt Sinn)
 - b) in der Empirie (Streudiagramm=Scatterplot)
3. Positive/negative Korrelation: ein höherer Wert bei Merkmal A geht mit einem höheren/niedrigeren Wert bei Merkmal B einher-
Achtung: Bevor Sie eine inhaltliche Interpretation abgeben:
Überprüfung der **Richtung** der Kodierung!!!
4. Gefahr von Scheinkorrelation / Redundanz:
kein kausaler Zusammenhang, aber gemeinsame Ursache, oder:
selbst wenn es eine Korrelation gibt, sagt das nichts darüber aus,
was die Ursache, und was die Wirkung ist
(SMV: sexueller Missbrauch und Trennung der Eltern)
Intervenierende Variablen abchecken!







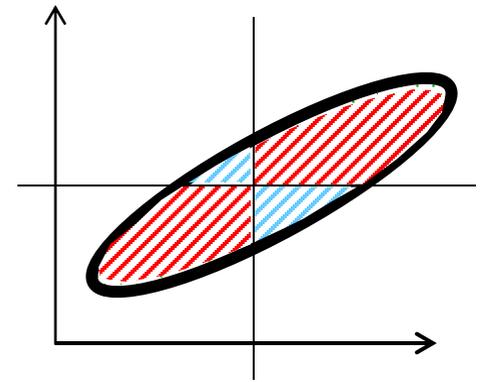
$r = -0,7$



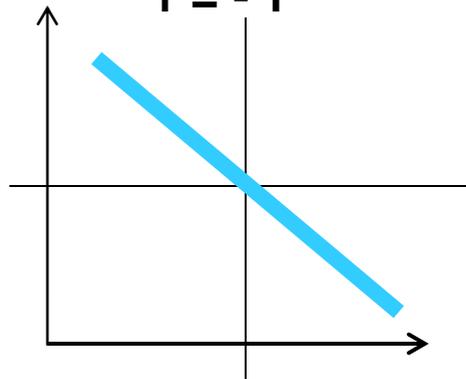
Korrelation im Bereich von:

-1 0 +1

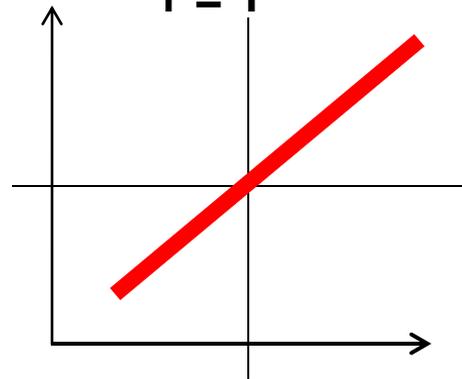
$r = 0,7$



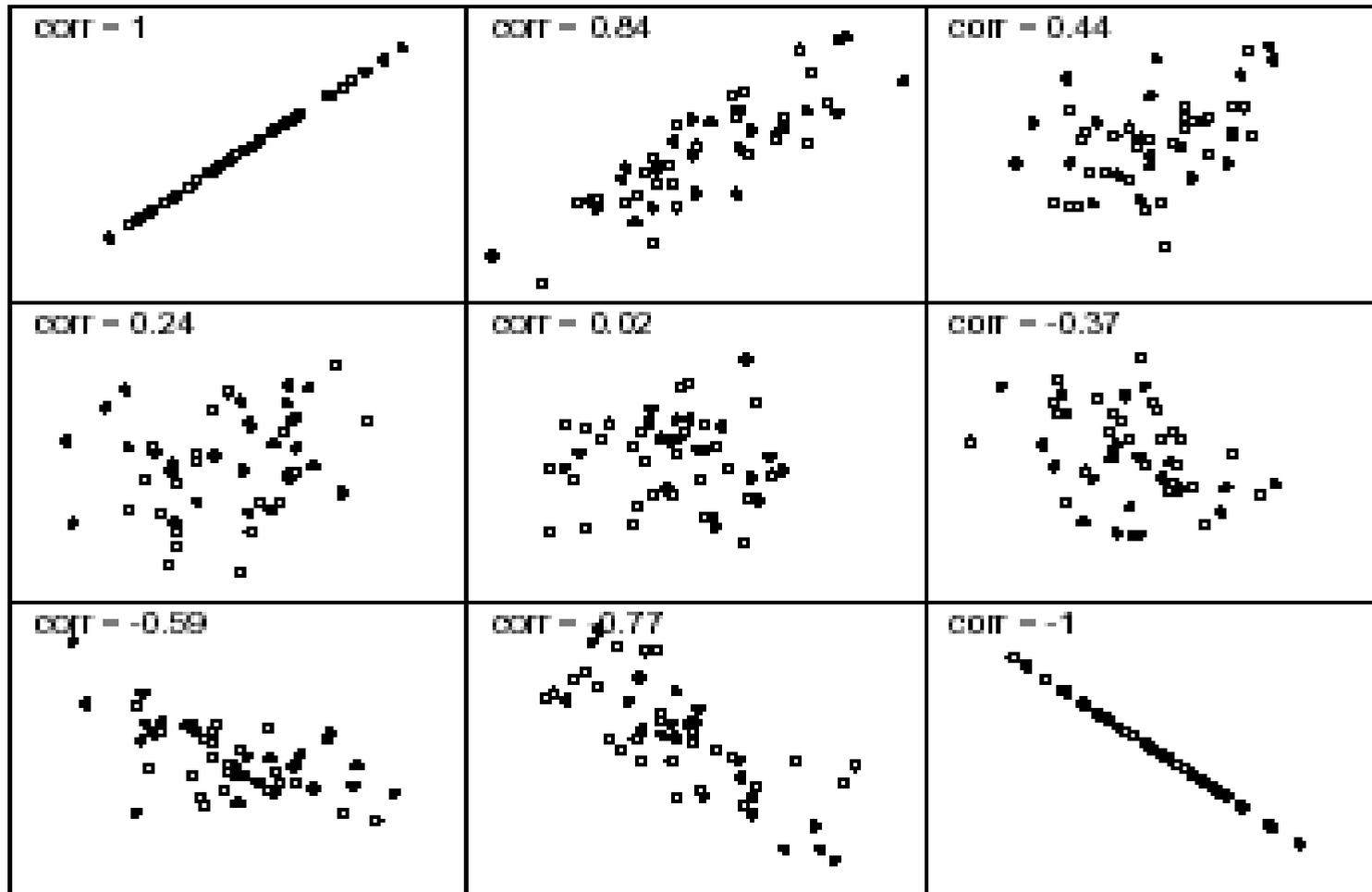
$r = -1$



$r = 1$



Punkteschwärme: Stärke der Korrelation



**Pearson-Korrelation von zwei metrischen Variablen:
Je länger das Gespräch dauert, desto mehr Wörter werden gesprochen.**

Korrelationen

		summ P A
dauer	Korrelation nach Pearson	,760**
	Signifikanz (2-seitig)	,000
	N	37

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig)

Korrelations-
koeffizient r

Signifikanz

Fallzahl

Spearman-Korrelation zwischen zwei ordinalen Variablen bzw. einer ordinalen und einer metrischen Variable:

Je stärker österreichisches Deutsch zur eigenen Identität gehört, desto mehr stören Begriffe wie „Blumenkohl“ oder „Brötchen“ beim Einkaufen.

		v19 Österreichisches Deutsch - das ist das, wo ich mich wohlfühle, es gehört zu meiner Identität	v22 Stören: Begriffe wie Blumenkohl oder Brötchen beim Einkaufen
Spearman-Rho	v19 Österreichisches Deutsch - das ist das, wo ich mich wohlfühle, es gehört zu meiner Identität	Korrelationskoeffizient	1,000
		Sig. (2-seitig)	.
		N	30
	v22 Stören: Begriffe wie Blumenkohl oder Brötchen beim Einkaufen	Korrelationskoeffizient	,291
		Sig. (2-seitig)	,119
		N	30

Signifikanz des Korrelationskoeffizienten
11,9% ist > 5% (nicht signifikant)

Spearman-Rho	v19 Österreichisches Deutsch - das ist das, wo ich mich wohlfühle, es gehört zu meiner Identität	
v22 Stören: Begriffe wie Blumenkohl oder Brötchen beim Einkaufen	Korr-Koeff.	,291*
	Sign.	,119
	n	30

ungewichtete Daten:
Korrelation nicht signifikant

Spearman-Rho	v19 Österreichisches Deutsch - das ist das, wo ich mich wohlfühle, es gehört zu meiner Identität	
v22 Stören: Begriffe wie Blumenkohl oder Brötchen beim Einkaufen	Korr-Koeff.	,291*
	Sign.	,024
	n	60

Fallzahl x 2:
Korrelation signifikant ($p < 0,05$)

Spearman-Rho	v19 Österreichisches Deutsch - das ist das, wo ich mich wohlfühle, es gehört zu meiner Identität	
v22 Stören: Begriffe wie Blumenkohl oder Brötchen beim Einkaufen	Korr-Koeff.	,291*
	Sign.	,005
	n	90

Fallzahl x 3:
Korrelation hochsignifikant ($p < 0,01$)
schwacher Zusammenhang

Spearman-Rho	v19 Österreichisches Deutsch - das ist das, wo ich mich wohlfühle, es gehört zu meiner Identität	
v22 Stören: Begriffe wie Blumenkohl oder Brötchen beim Einkaufen	Korr-Koeff.	,291*
	Sign.	,000
	n	300

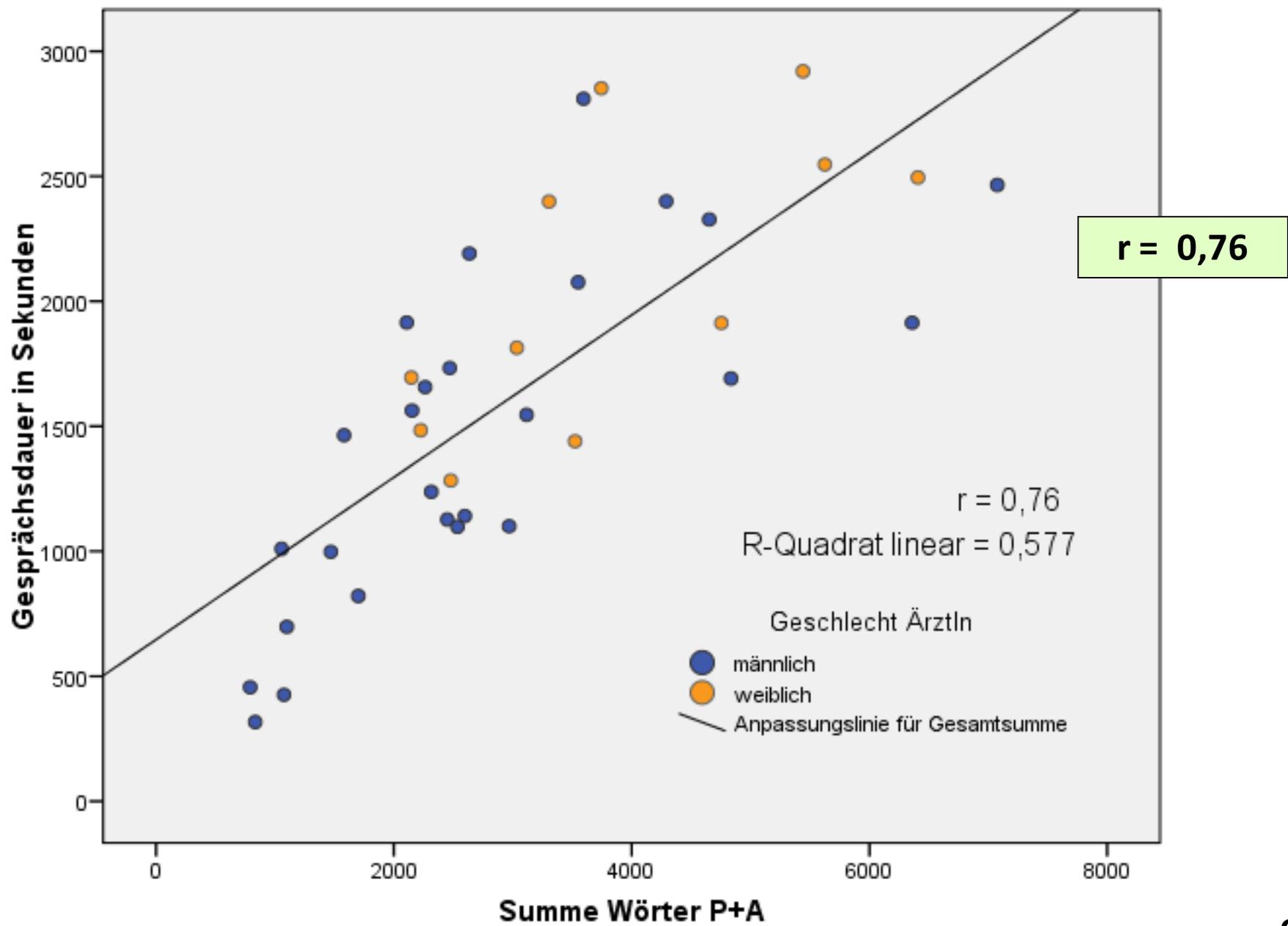
Fallzahl x 10:
Korrelation hochsignifikant ($p < 0,01$)
mittlerer Zusammenhang

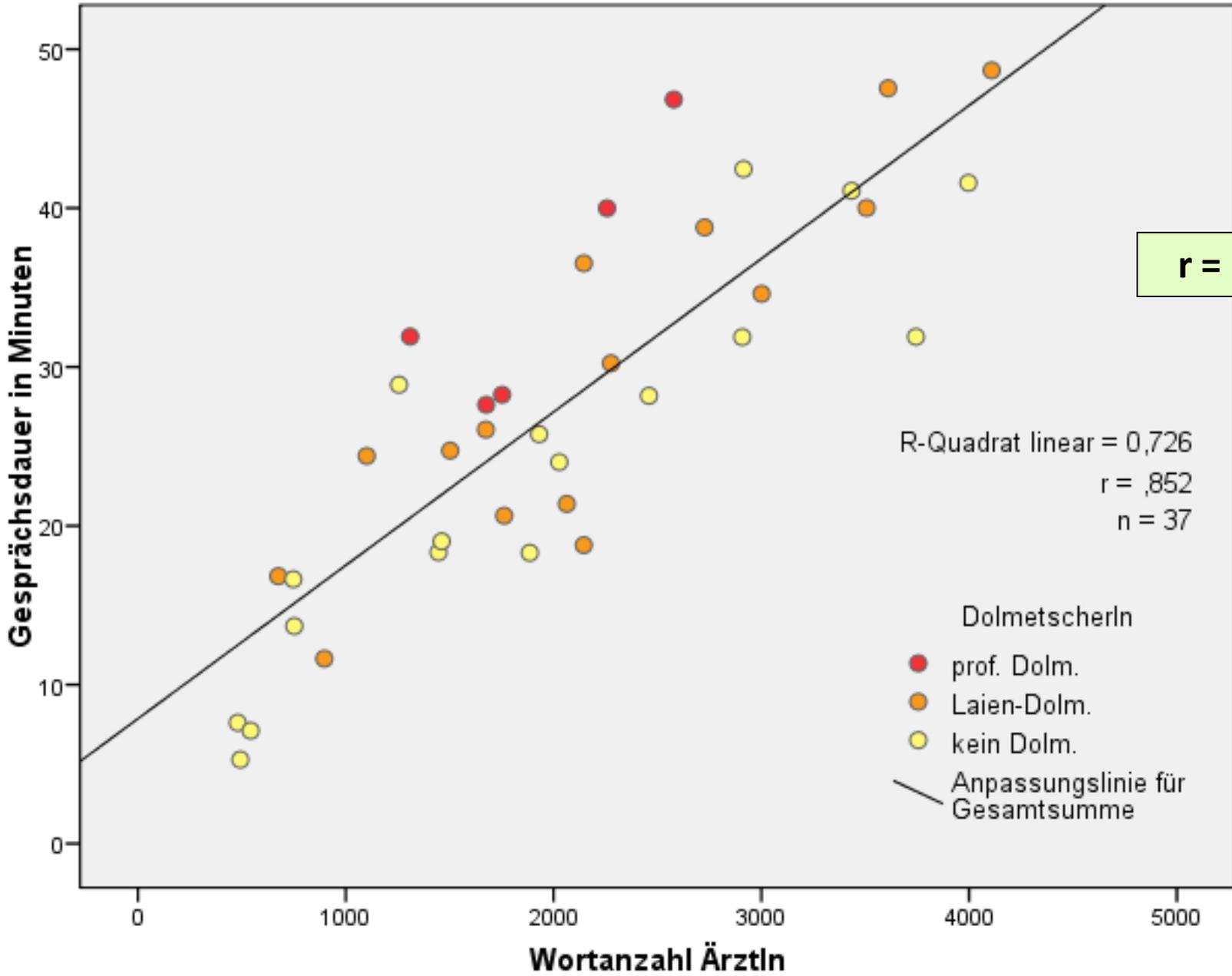
Korrelation und Signifikanz

Zu jedem Prüfmaß bei einem Signifikanztest gibt es eine Signifikanz.
 Auch zu jedem Korrelationskoeffizienten gibt es eine Signifikanz. Wieder gilt:
 Je höher die Fallzahl, desto schwächere Korrelationen sind bereits signifikant.

n	r ab dem die Korrelation signifikant ist	schwacher Zushg.	mittlerer Zushg.	starker Zushg.
5	0,75			
10	0,58			
25	0,40	0,5	0,6	0,7
50	0,27	0,4	0,5	0,6
100	0,2	0,3	0,4	0,5
500	0,09	0,2	0,3	0,4
1000	0,06	0,15	0,25	0,35
2000	0,04	0,1	0,2	0,3

Grau: Achtung:
 Keine Konvention!!
 Erfahrungswerte
 von I.H.

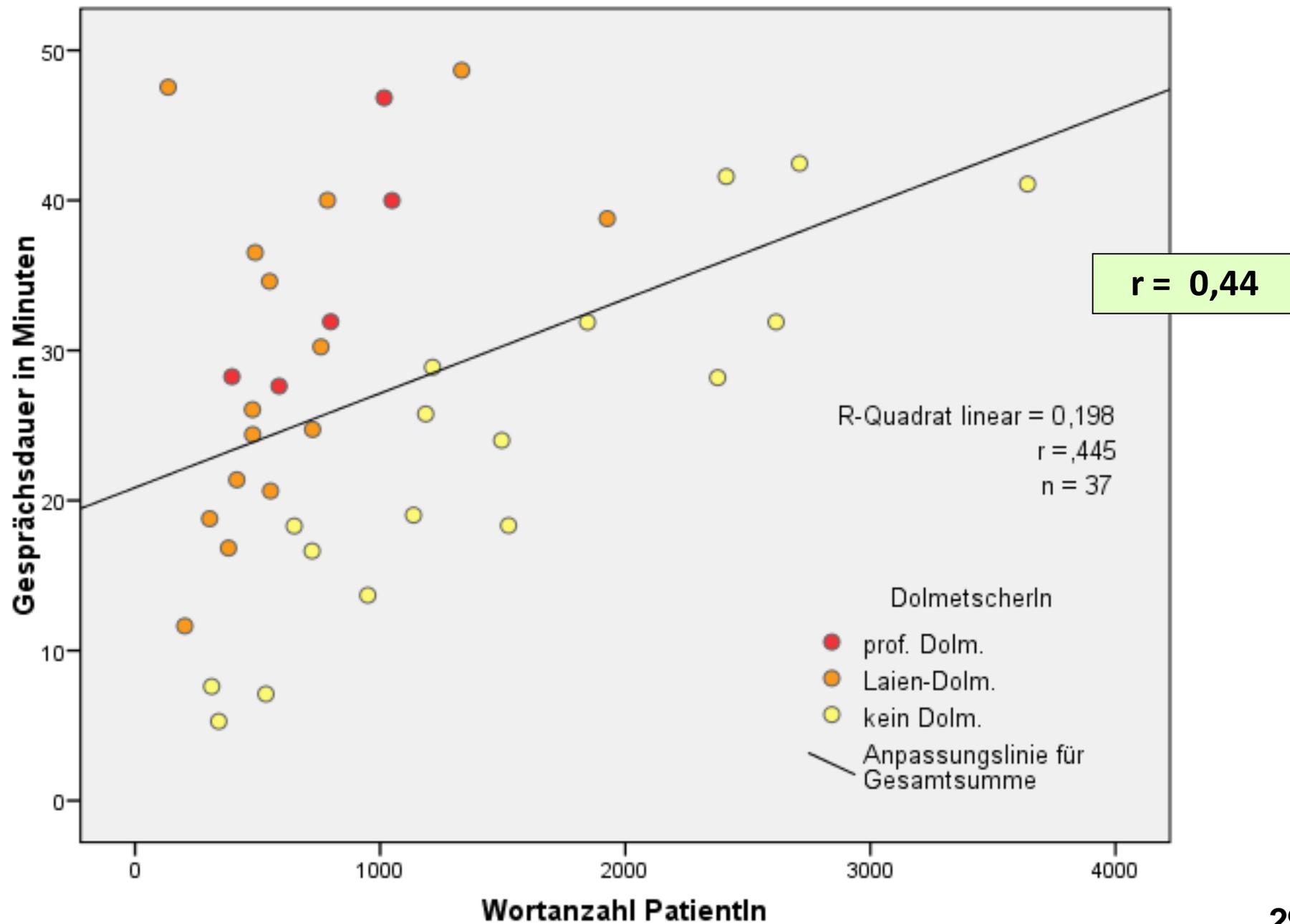


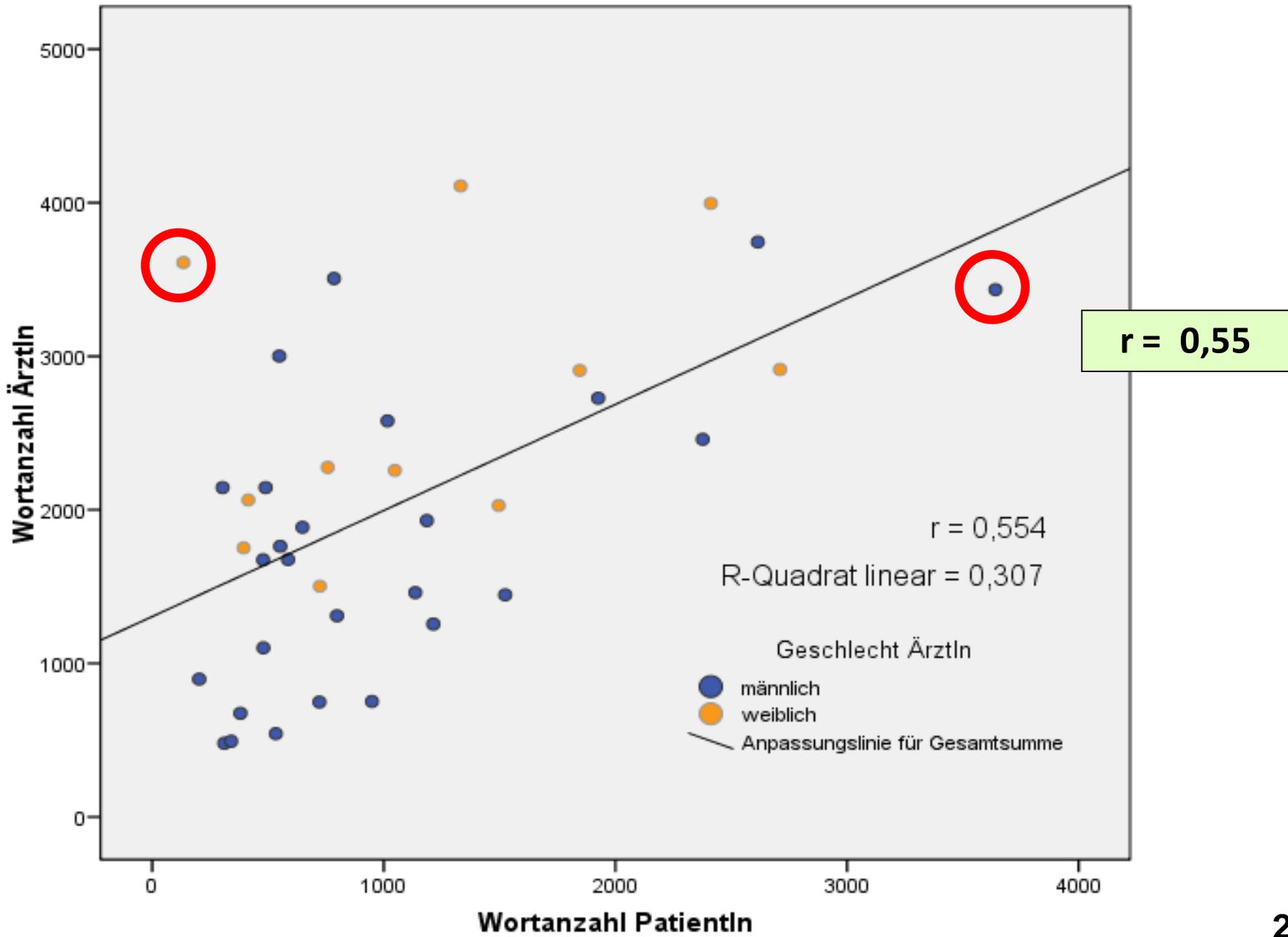


r = 0,85

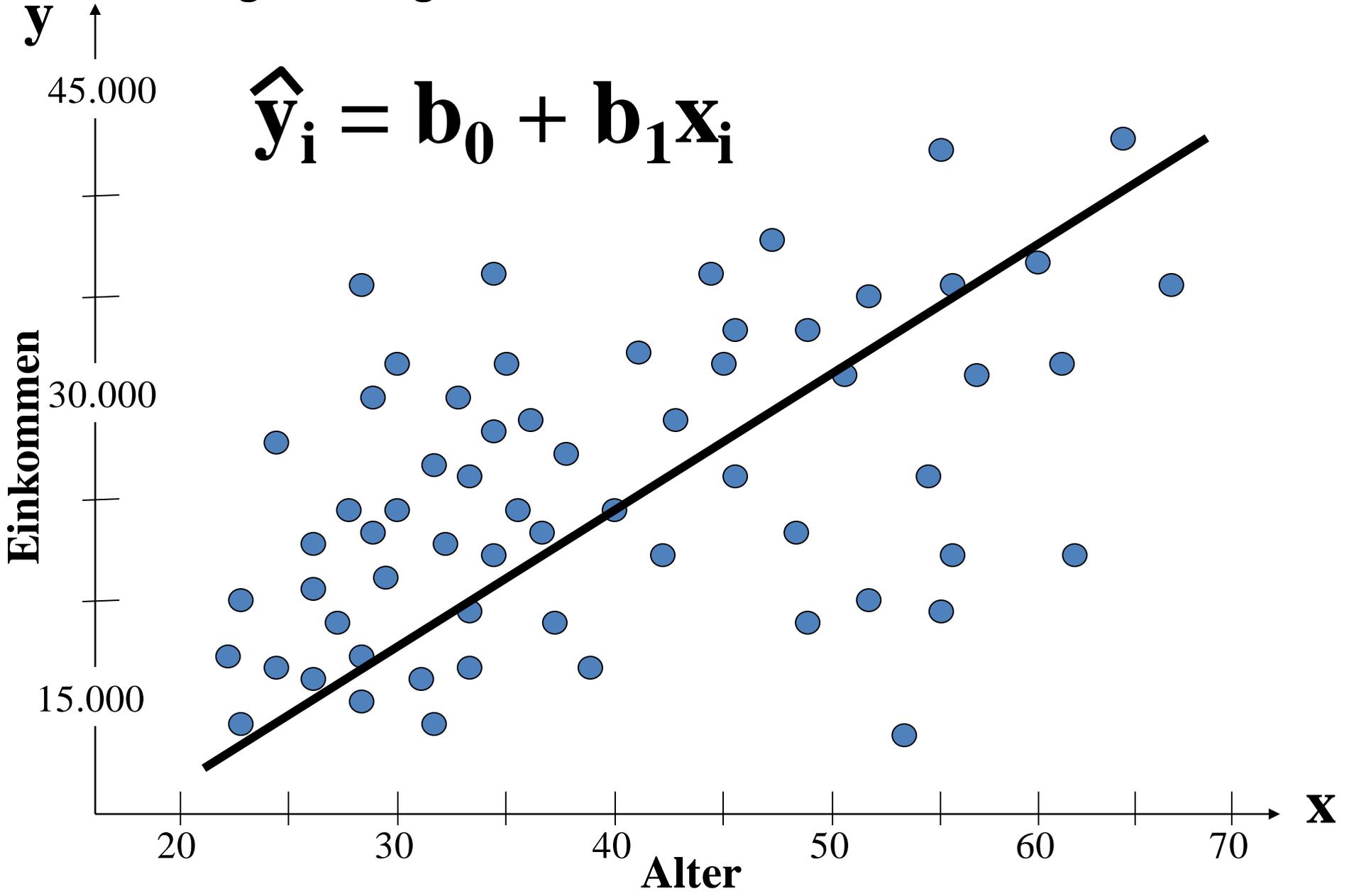
R-Quadrat linear = 0,726
r = ,852
n = 37

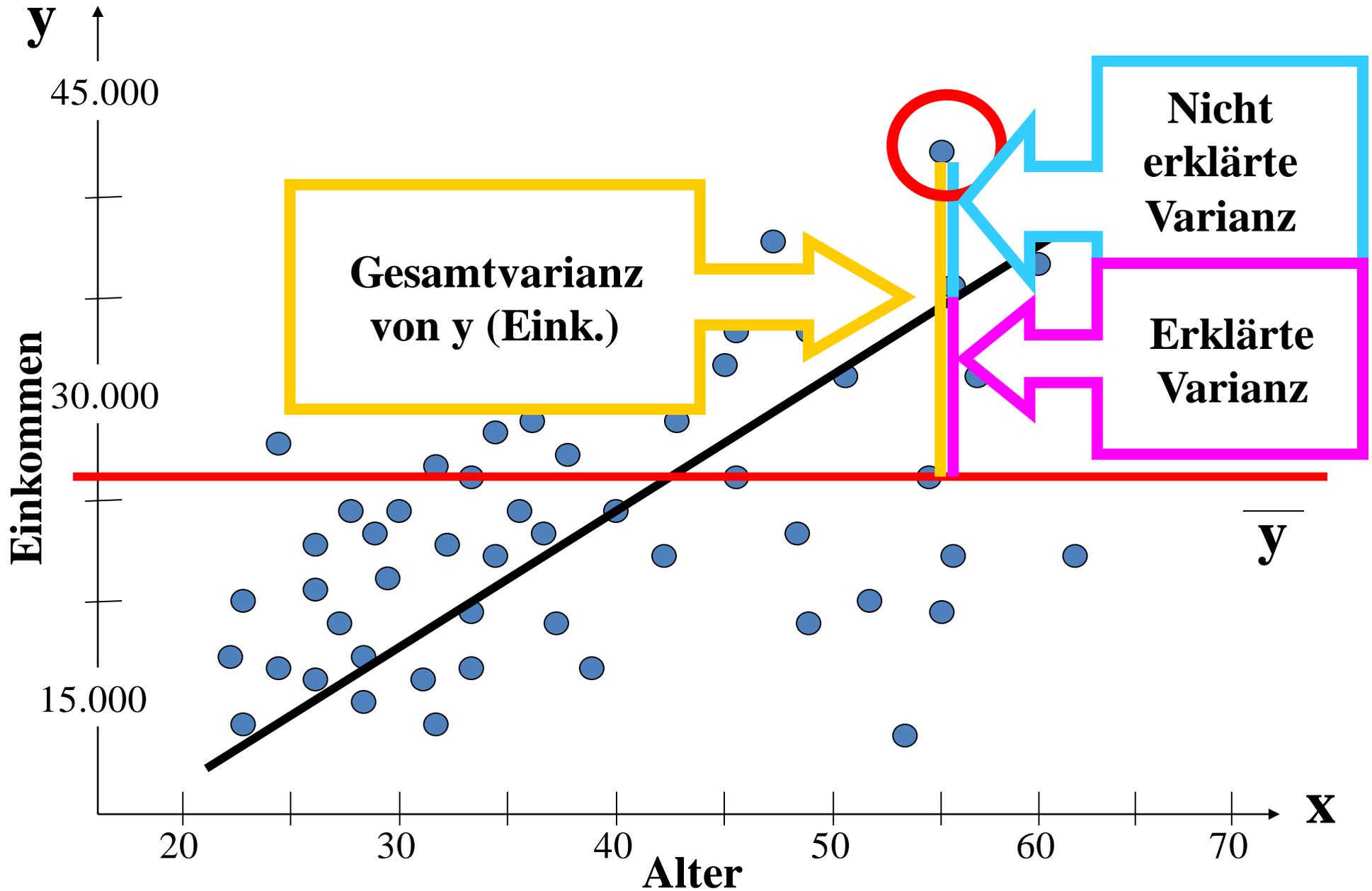
- DolmetscherIn
- prof. Dolm.
- Laien-Dolm.
- kein Dolm.
- Anpassungslinie für Gesamtsumme

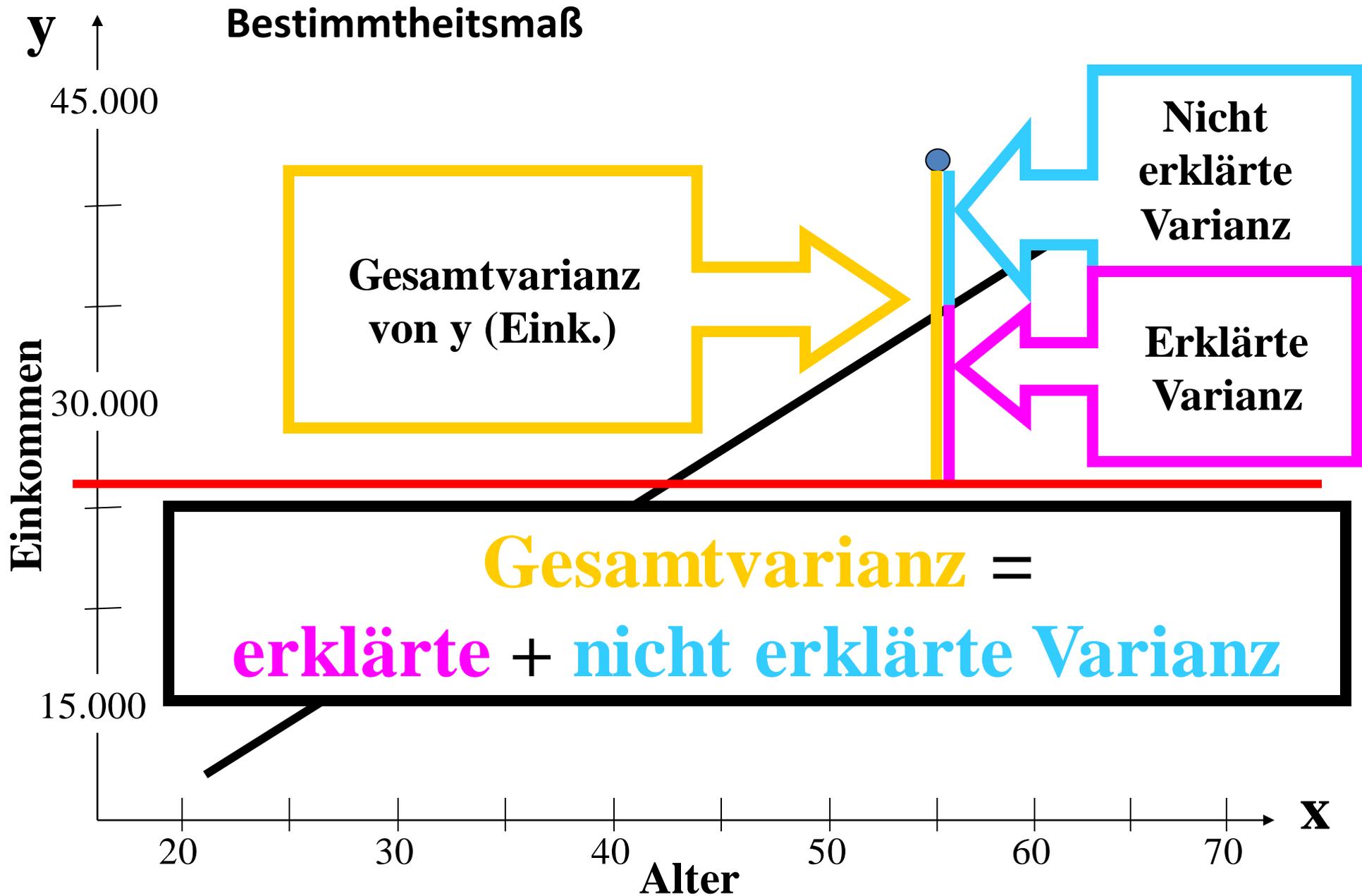


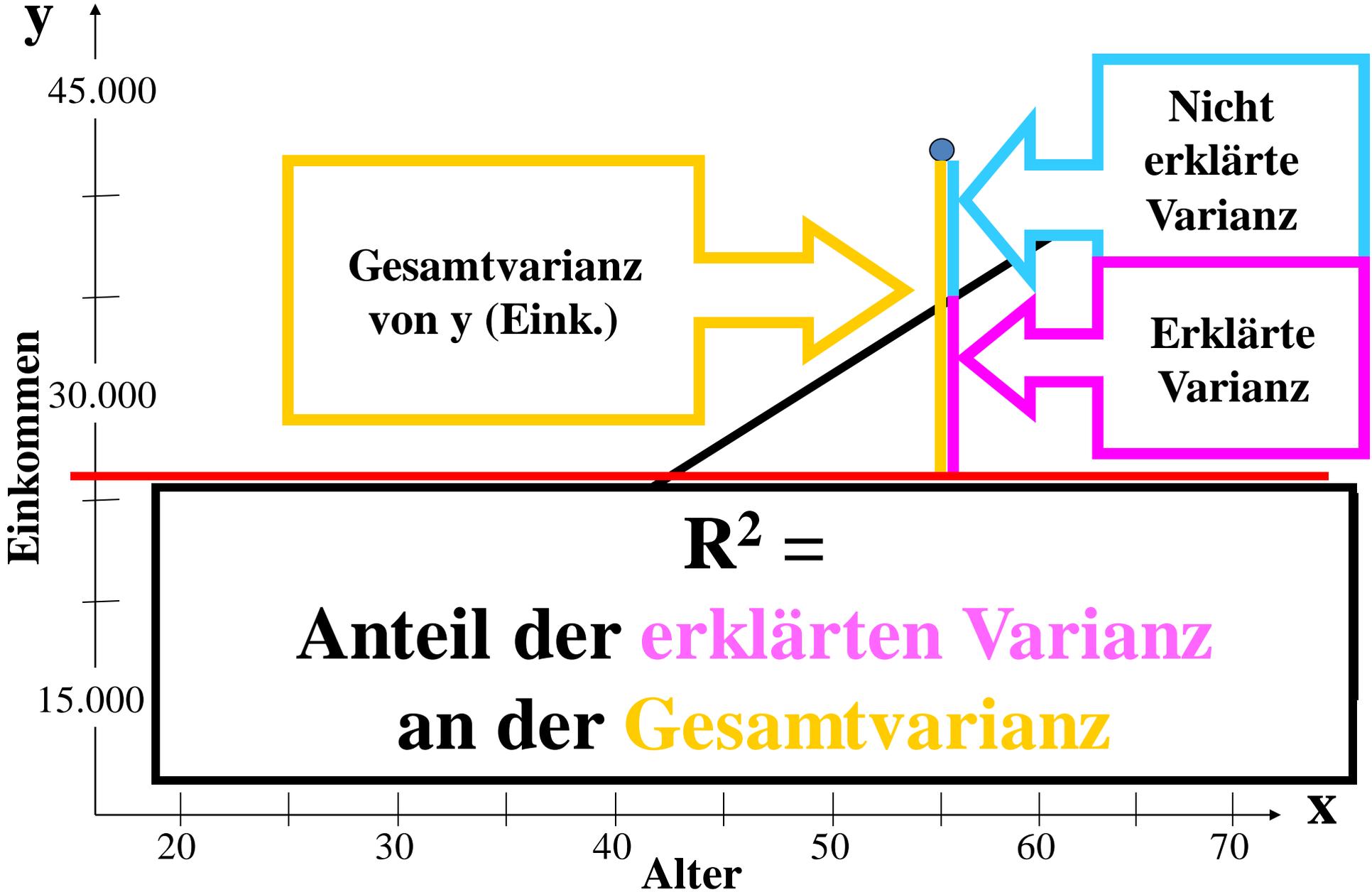


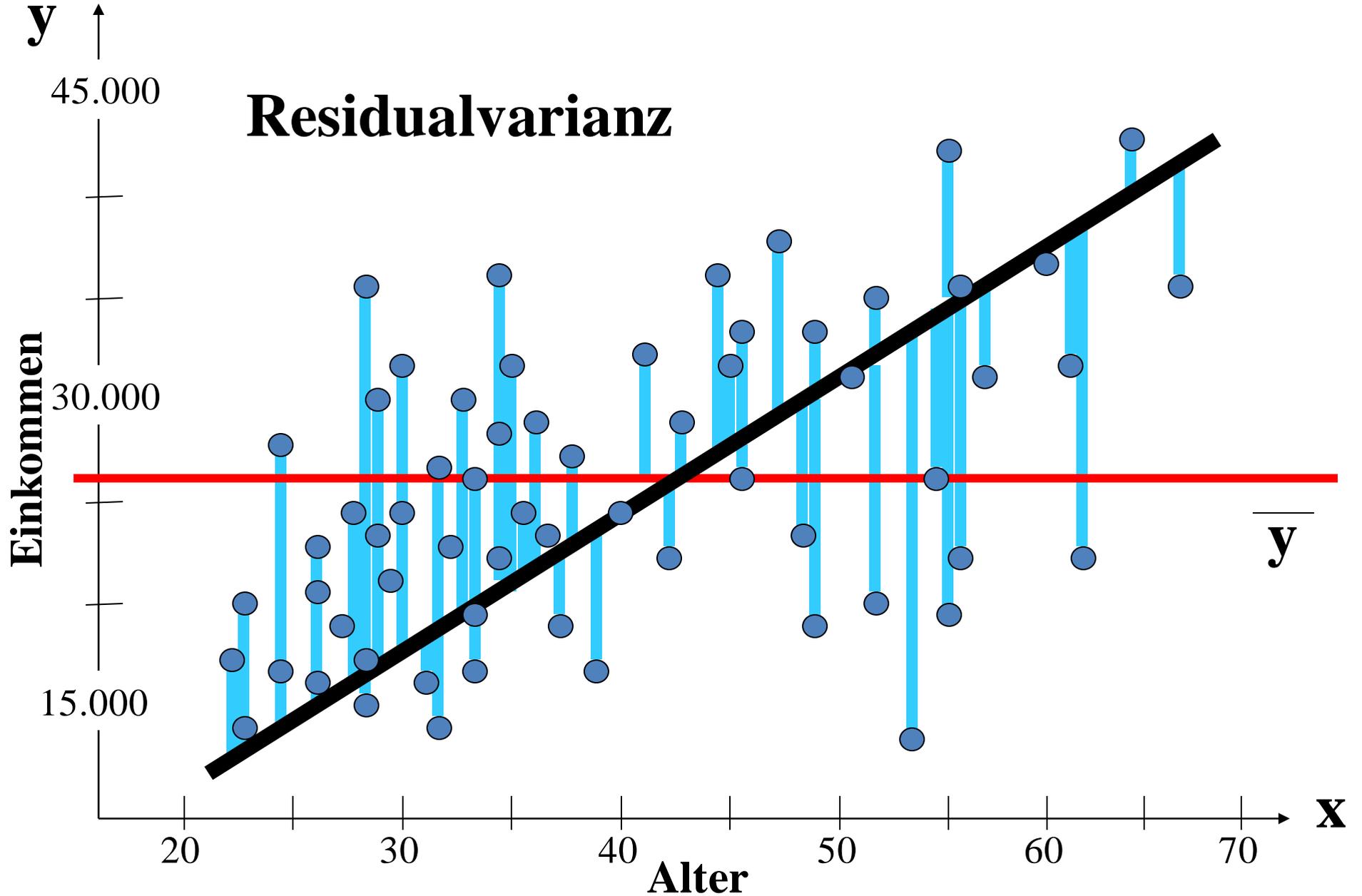
Regressionsgerade

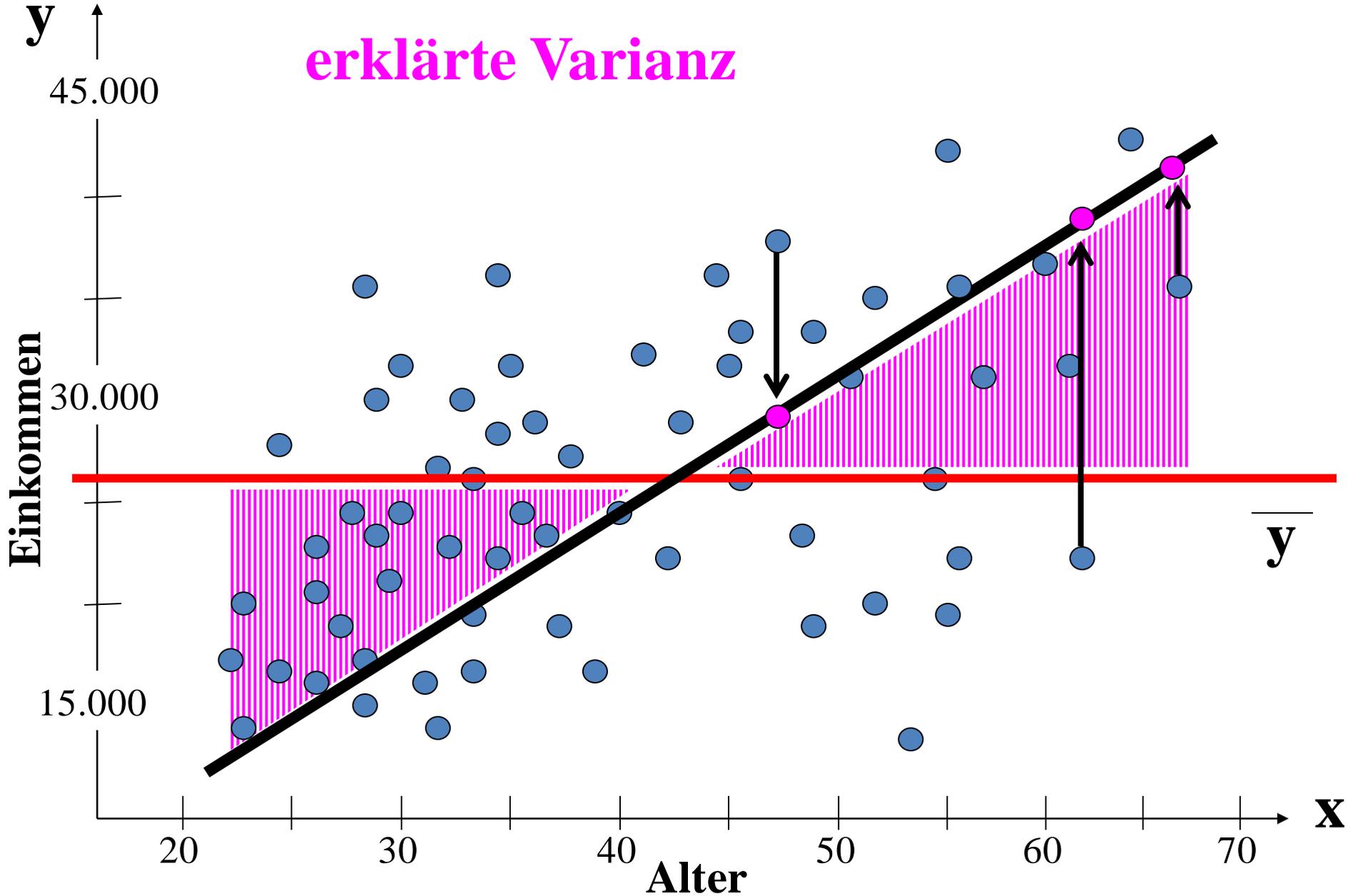










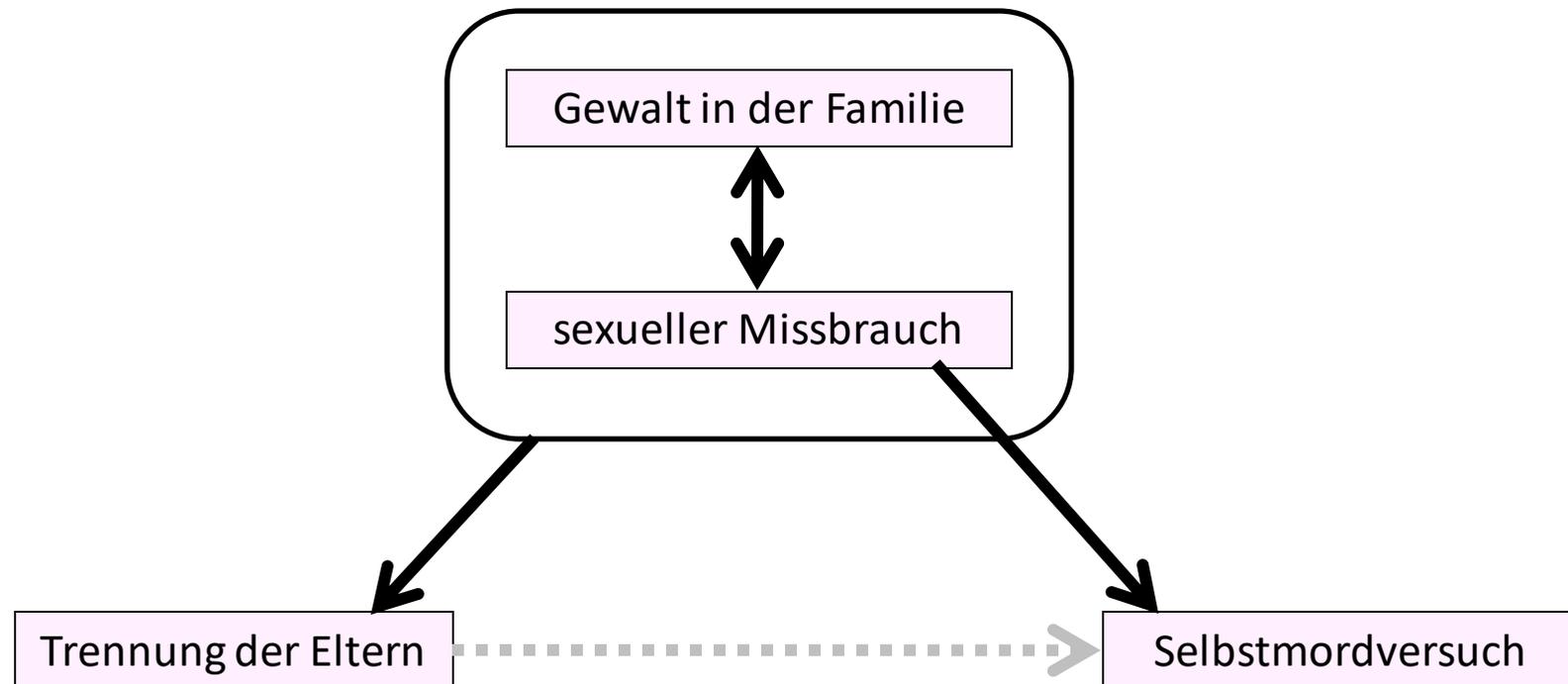


Korrelation gibt keine Auskunft über Kausalität

Vorsicht vor Scheinkorrelationen.

Wichtig: Bezugnahme auf ein theoretisches Modell der Gesetzmäßigkeiten.

Mehrere Variablen in die Analyse miteinbeziehen.



„Im Bett ist es gefährlich! – Im Bett sterben die meisten Menschen.“

Siehe kuriose Scheinkorrelationen unter: <http://www.tylervigen.com/spurious-correlations>

Indexberechnung Dimensionierung

Indexberechnung = Information aus mehreren Items wird in einer Variable zusammengefasst

Indexberechnung mittels COUNT-Befehl: **ZÄHL-INDEX**

z.B. Index „Anzahl besonderer Ereignisse“ aus 11 Items-Mehrfachantwort zu „lifeevents“ (0 = genannt / 1 = nicht genannt)

→ aus mehreren dichotomen Variablen wird eine metrische Variable.

Indexberechnung mittels COMPUTE-Befehl: **SUMMEN-INDEX**

z.B. Index „Allgemeines Wohlbefinden“ aus den Variablen

(1 „very well“ bis 5 „very bad“)

g14 Contentedness with life

g15 your physical condition

g16 your emotional / mental health

→ aus mehreren ordinalen Variablen wird eine metrische Variable.

Wozu: Bei umfangreichen Erhebungsinstrumenten werden Einzelitems, die laut theoretischem Modell „das gleiche“ messen, in **Messdimensionen (= Indizes)** zusammengefasst.

Erleichtert: korrelative und multivariate Verfahren, Erkenntnisgewinn.

Mehrfachantworttabelle

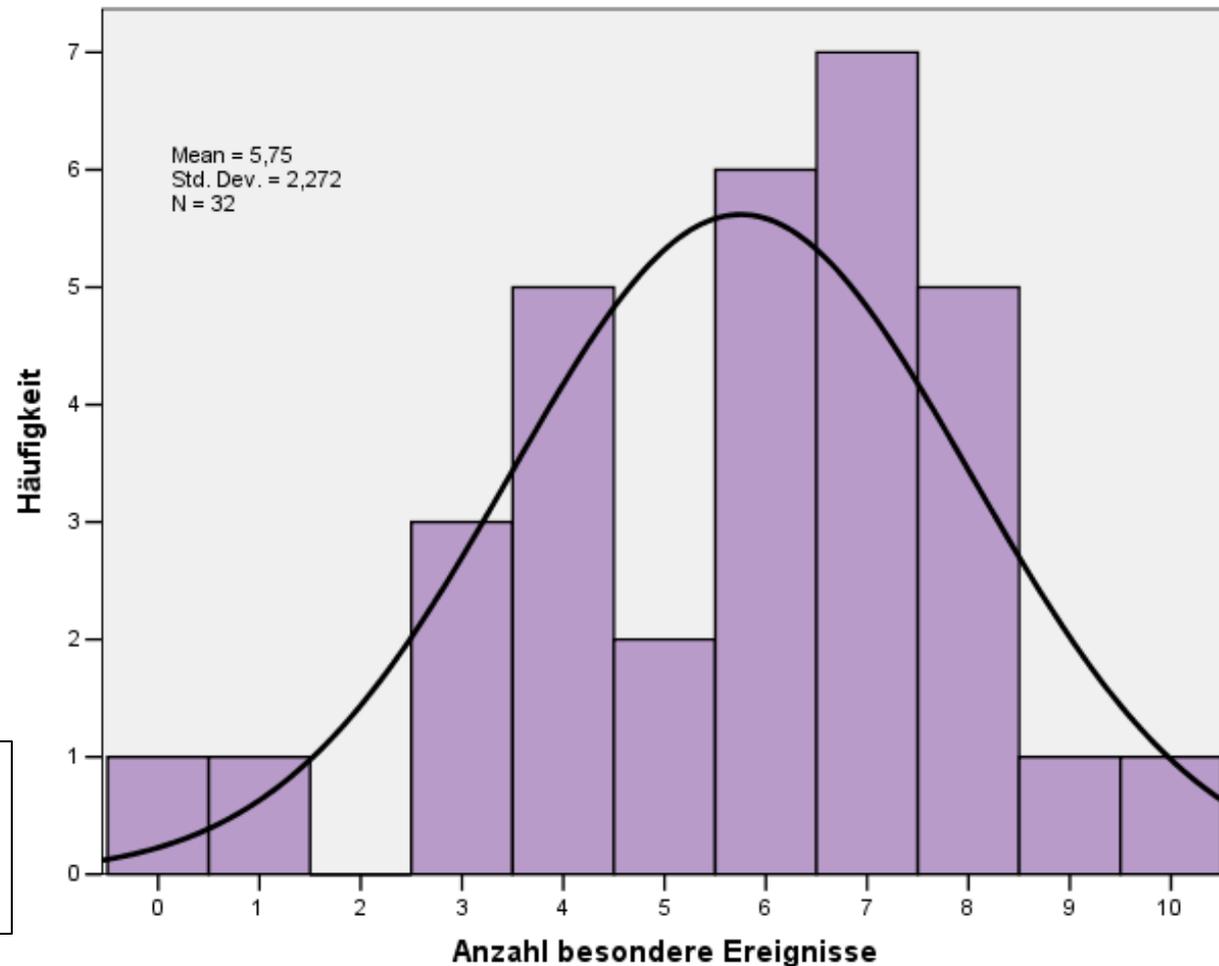
	Anzahl	Anzahl als Spalten (%)
b1 separation or divorce of parents	22	68,8%
b2 death of mother or father	10	31,3%
b3 alcohol, medicament or drug abuse of father or mother	16	50,0%
b4 grow up in a foster family	5	15,6%
b5 physical violence or threats from family members, relatives or known persons	15	46,9%
b6 sexual harassments or sexual abuse from family members, relatives or known persons	14	43,8%
b7 runaway from home as child or adolescent because of conflicts	19	59,4%
b8 death of a close girlfriend or boy friend	24	75,0%
b9 serious sickness of a close person	17	53,1%
b10 own serious sickness	24	75,0%
b11 attempted suicide	18	56,3%
b0 none	1	3,1%
Gesamt	32	100,0%

Zählindex = Indexberechnung mittels „COUNT“

anz_lifevents Anzahl besondere Ereignisse

		Anz.	%	Kum. %
Gültig	0	1	3,1	3,1
	1	1	3,1	6,3
	3	3	9,4	15,6
	4	5	15,6	31,3
	5	2	6,3	37,5
	6	6	18,8	56,3
	7	7	21,9	78,1
	8	5	15,6	93,8
	9	1	3,1	96,9
	10	1	3,1	100,0
Ges.		32	100,0	

Zur Darstellung mittels Kreuztabellen
oder Balkendiagrammen:
Umkodierung in Kategorien.



Summenindex = Indexberechnung mit „COMPUTE“

	g14	g15	g16
	Contentedness with life	your physical condition	your emotional / mental health
	%	%	%
1 very well	9,4%	15,6%	9,4%
2 fine	31,3%	34,4%	15,6%
3 mean	37,5%	28,1%	37,5%
4 bad	9,4%	12,5%	21,9%
5 very bad	12,5%	9,4%	15,6%
Total	100,0%	100,0%	100,0%

Summieren: Summe der Werte in den Variablen $g14 + g15 + g16$

Index ohne Normierung

```
COMPUTE sum_wohlbefinden =
g14 + g15 + g16 .
```

sum_wohlbefindenxx Index Allgemeines Wohl

		Häufigkeit	Prozent
Gültig	4	1	3,1
	5	2	6,3
	6	4	12,5
	7	5	15,6
	8	7	21,9
	9	4	12,5
	10	2	6,3
	11	2	6,3
	13	2	6,3
	15	3	9,4
	Gesamt	32	100,0

Skala 3 -15

Index mit Normierung

```
COMPUTE sum_wohlbefinden =
(((g14 + g15 + g16) - 3) / 12) * 100 .
```

sum_wohlbefinden Index Allgemeines Wohlbefinden

		Häufigkeit	Prozent	Kum. %
Gültig	8	1	3,1	3,1
	17	2	6,3	9,4
	25	4	12,5	21,9
	33	5	15,6	37,5
	42	7	21,9	59,4
	50	4	12,5	71,9
	58	2	6,3	78,1
	67	2	6,3	84,4
	83	2	6,3	90,6
	100	3	9,4	100,0
	Gesamt	32	100,0	

Skala 0 -100

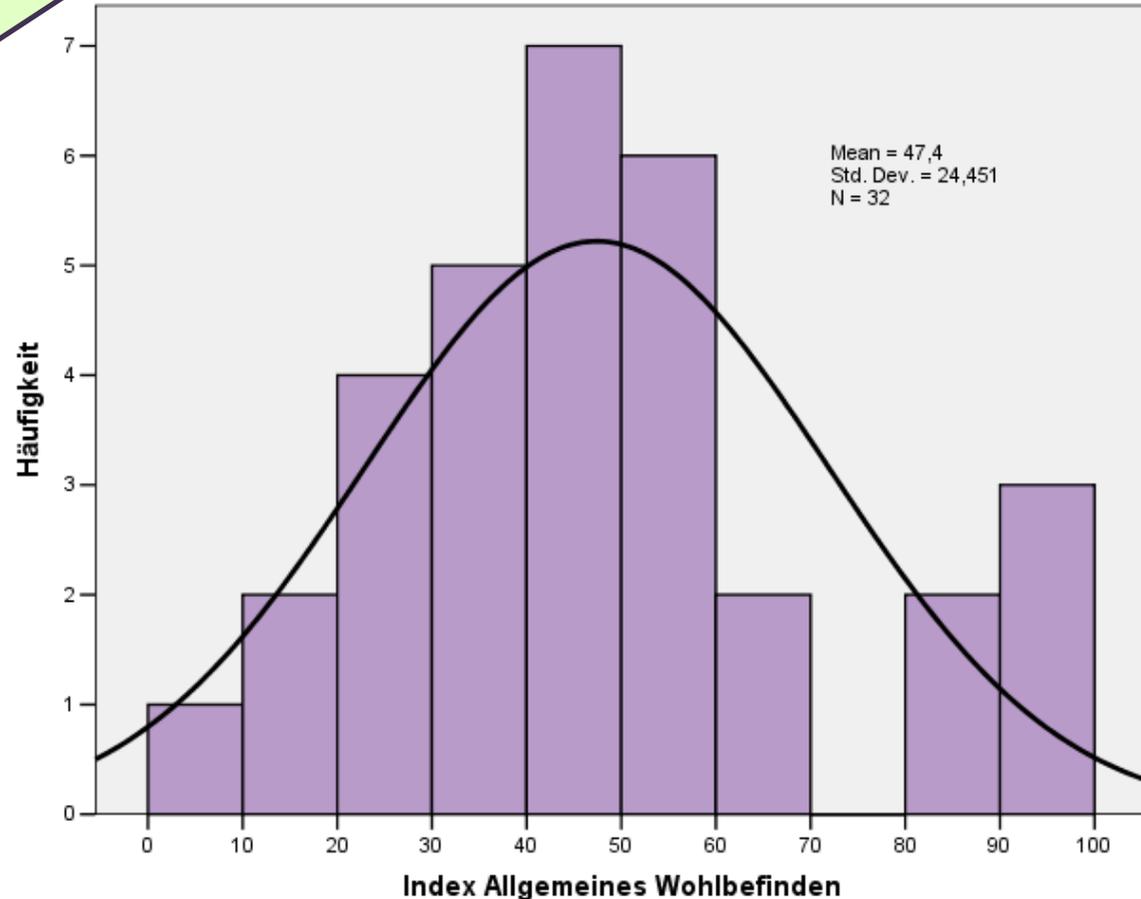
Index-Normierung

```
COMPUTE sum_wohlbe finden = (((g14 + g15 + g16) - 3) / 12) * 100 .
```

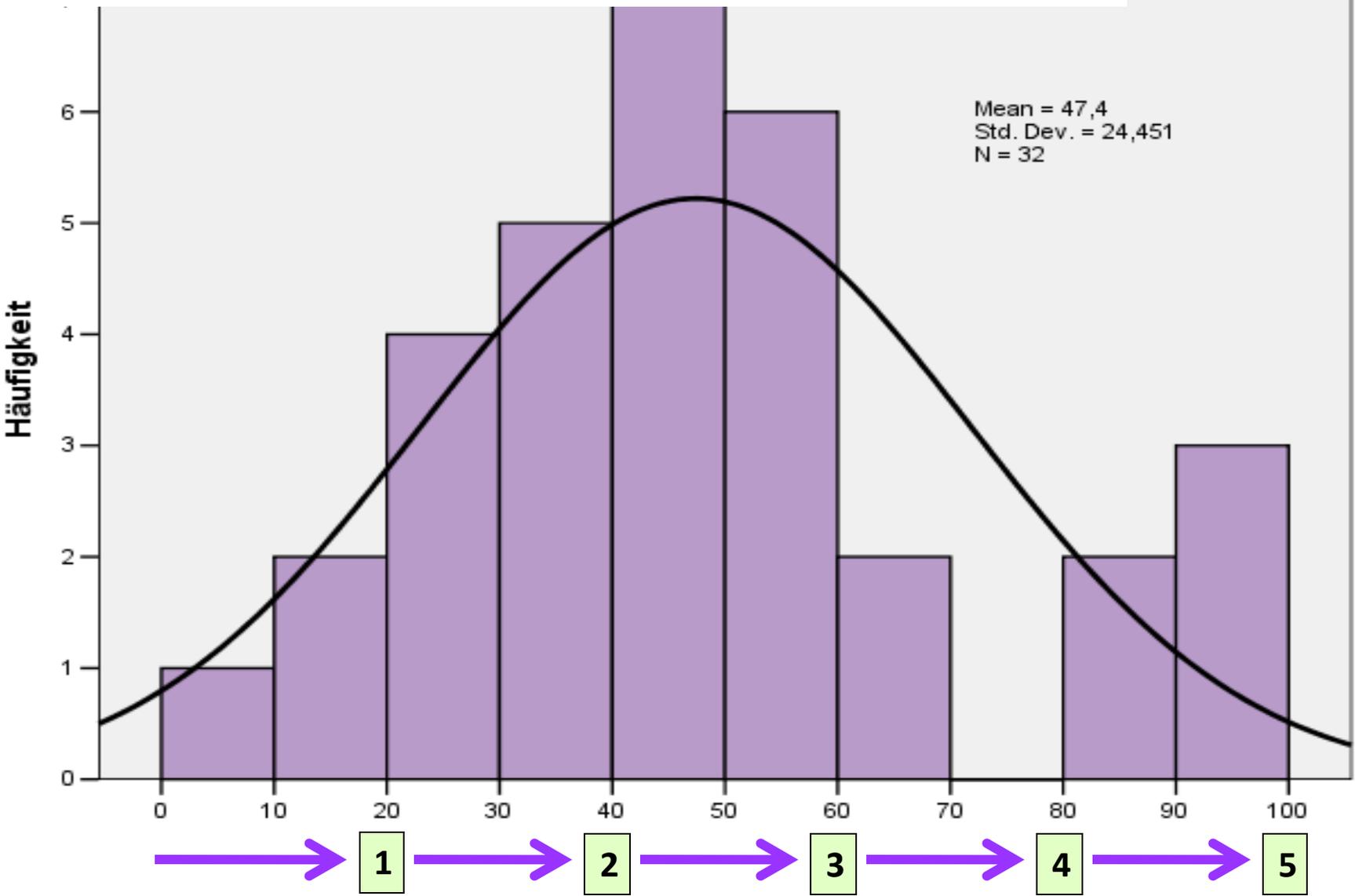
Skala 3-15: kleinsten Wert abziehen
(= Verankern im Nullpunkt)

Skala 0-12: durch größten Wert
(= Normieren auf 1)

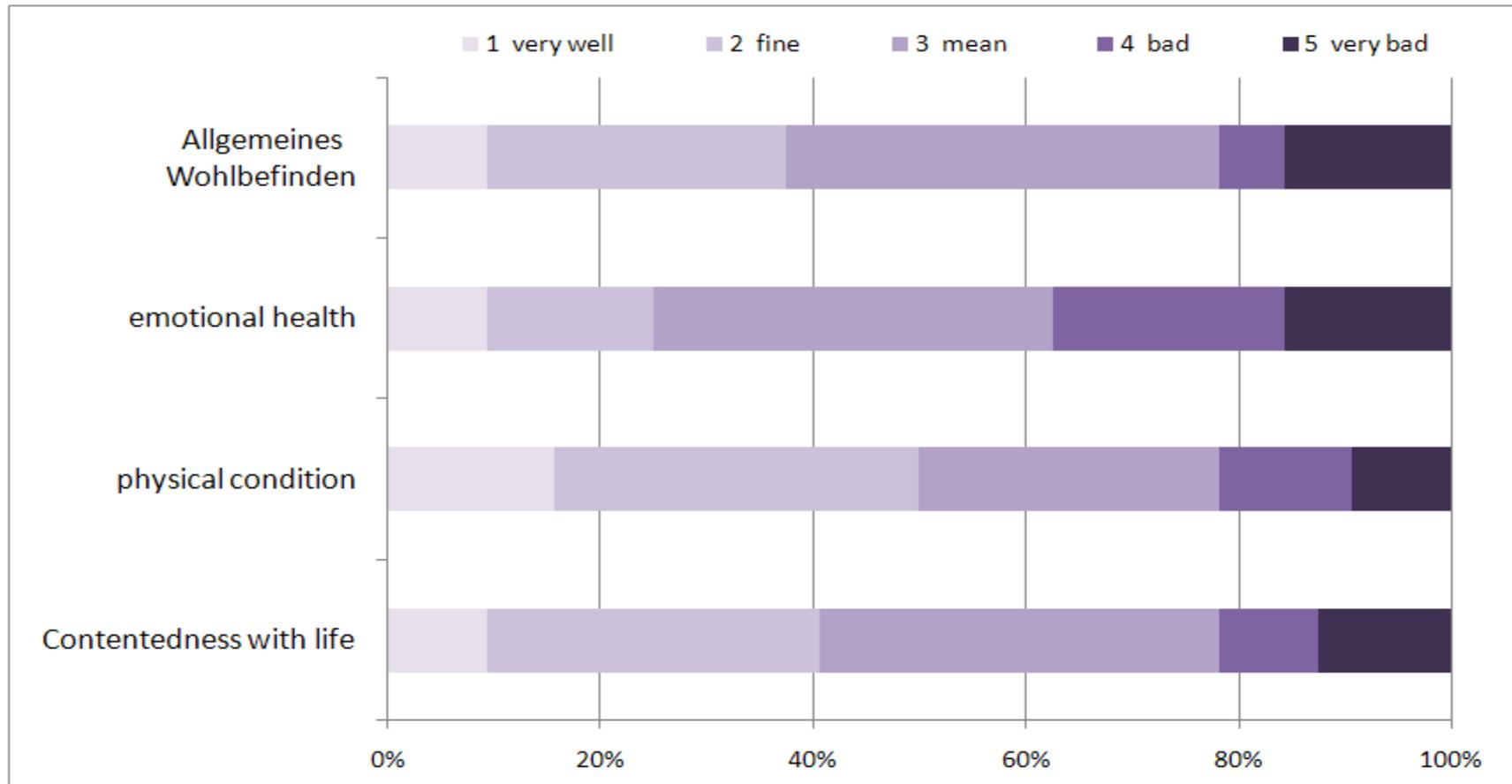
Skala 0-1: * 100 (= Prozentuieren)



Rückführung des Index in die Originalskala



	Contentedness with life	your physical condition	your emotional / mental health	Allgemeines Wohlbefinden
	%	%	%	%
1 very well	9,4%	15,6%	9,4%	9,4%
2 fine	31,3%	34,4%	15,6%	28,1%
3 mean	37,5%	28,1%	37,5%	40,6%
4 bad	9,4%	12,5%	21,9%	6,3%
5 very bad	12,5%	9,4%	15,6%	15,6%
Total	100,0%	100,0%	100,0%	100,0%



Korrelationsmatrix

Korrelationsmatrix

	COL A	KAF F	OBS T	SUES S	GEM US	PENU T	CHIP S	FRI T	HAM B	VOLL K	MILK	YOG
COLA Wie oft Coca Cola	1,000	,120	-,037	,293	-,108	,143	,340	,337	,292	-,145	-,053	-,052
KAFF Wie oft Kaffee	,120	1,00	-,047	,068	-,010	,050	,072	,063	,063	-,025	,045	-,023
OBST Wie oft Obst	-,037	-,047	1,000	-,014	,378	,133	,061	,047	-,017	,265	,144	,293
SUESS Wie oft Suessigkeiten	,293	,068	-,014	1,000	-,069	,114	,294	,217	,213	-,128	-,056	-,007
GEMUSE Wie oft Gemuese	-,108	-,010	,378	-,069	1,000	,140	-,017	-,030	-,020	,309	,201	,283
PENUT Wie oft Erdnuesse	,143	,050	,133	,114	,140	1,000	,379	,334	,226	,114	,058	,111
CHIPS Wie oft Chips	,340	,072	,061	,294	-,017	,379	1,000	,455	,336	-,052	-,039	,003
FRIT Wie oft Pommes frites	,337	,063	,047	,217	-,030	,334	,455	1,00	,463	-,031	,015	,004
HAMB Wie oft Hamburger	,292	,063	-,017	,213	-,020	,226	,336	,463	1,00	-,029	,033	,016
VOLLK Wie oft Vollkornbrot	-,145	-,025	,265	-,128	,309	,114	-,052	-,031	-,029	1,000	,225	,343
MILK Wie oft Milch	-,053	,045	,144	-,056	,201	,058	-,039	,015	,033	,225	1,000	,323
YOG Wie oft Topfen Yoghurt	-,052	-,023	,293	-,007	,283	,111	,003	,004	,016	,343	,323	1,00

Ergebnis aus einer Faktoranalyse

Rotierte Komponentenmatrix^a

	Komponente		
	1	2	3
FRIT Wie oft Pommes frites	,758		
CHIPS Wie oft Chips	,755		
HAMB Wie oft Hamburger	,658		
COLA Wie oft Coca Cola	,606		
PENUT Wie oft Erdnuesse	,558	,302	
SUESS Wie oft Suessigkeiten	,502		
GEMUSE Wie oft Gemuese		,681	
YOG Wie oft Topfen Yoghurt		,679	
VOLLK Wie oft Vollkornbrot		,669	
OBST Wie oft Obst		,644	
MILK Wie oft Milch		,510	,472
KAFF Wie oft Kaffee			,811

← „fast food“

← „gesund“

← „Kaffee“

Extraktionsmethode: Hauptkomponentenanalyse.
 Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 4 Iterationen konvergiert.

HBSC Gesundheitsrelevantes Verhalten von Schüler:innen, 1990
 n= 3206

Rotierte Komponentenmatrix^a

	Komponente		
	1	2	3
RTE10 Erdogan: gerecht - ungerecht	,879		
RTE8 Erdogan: ehrlich - unehrlich	,858		
RTE1 Erdogan: sympathisch - unsympathisch	,841		
RTE17 Erdogan: gut - böse	,727		
RTE6 Erdogan: für die Armen da - für die Reichen da	,718		
RTE5 Erdogan: gefährlich - ungefährlich	-,696	,327	
RTE12 Erdogan: korrupt - nicht korrupt	-,567		-,393
RTE11 Erdogan: verantwortungsvoll - verantwortungslos	,512		,494
RTE2 Erdogan: stark - schwach		,821	
RTE16 Erdogan: mächtig - ohnmächtig		,754	
RTE4 Erdogan: abhängig - unabhängig		,677	
RTE13 Erdogan: autoritär - demokratisch		,592	
RTE18 Erdogan: berechenbar - unberechenbar			,819
RTE14 Erdogan: emotional - rational			-,751
RTE9 Erdogan: ziellos - zielorientiert			-,758
RTE15 Erdogan: klug - dumm			,618
RTE3 Erdogan: koordiniert - chaotisch			,590
RTE7 Erdogan: kompetent - inkompetent			,503



Extraktionsmethode: Hauptkomponentenanalyse.
 Rotationsmethode: Varimax mit Kaiser-Normalisierung.

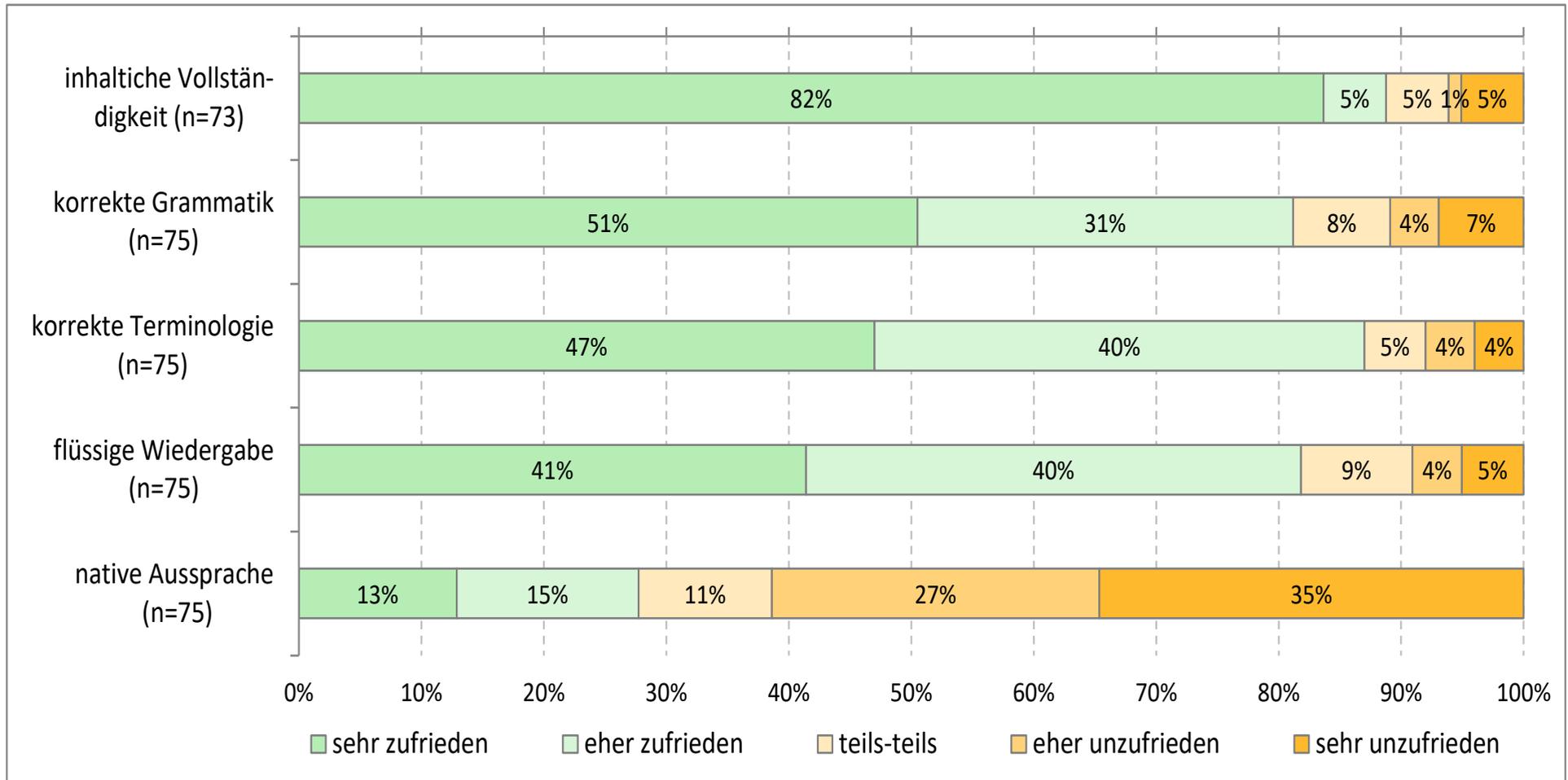
a. Die Rotation ist in 8 Iterationen konvergiert.

Überprüfen der Eindimensionalität einer Skala - Reliabilitätsanalyse

Bsp: Wichtigkeit verschiedener Aspekte für die Qualität von Simultanübersetzungen für TeilnehmerInnen an Tagungen/Kongressen

Wichtig für die Qualität von Simultanübersetzungen	5 sehr zufrieden		4 eher zufrieden		3 teils-teils		2 eher unzufrieden		1 sehr unzufrieden		Gesamt	
	Anz.	%	Anz.	%	Anz.	%	Anz.	%	Anz.	%	Anz.	%
	d Inhaltliche Vollständigkeit	60	82%	4	5%	4	5%	1	1%	4	5%	73
e Korrekte Grammatik	38	51%	23	31%	6	8%	3	4%	5	7%	75	100%
a Korrekte Terminologie	35	47%	30	40%	4	5%	3	4%	3	4%	75	100%
b Flüssige Wiedergabe	31	41%	30	40%	7	9%	3	4%	4	5%	75	100%
c Native Aussprache	10	13%	11	15%	8	11%	20	27%	26	35%	75	100%

Überprüfen der Eindimensionalität einer Skala - Reliabilitätsanalyse



Überprüfen der Eindimensionalität einer Skala - Reliabilitätsanalyse

Berechnen der Itemschwierigkeit

$$P_i = \frac{\sum x_i}{x_{max} * n} = \frac{\bar{x}}{x_{max}} \dots \text{wenn die Skala mit 0 beginnt.}$$

$$P_i = \frac{\bar{x} - x_{min}}{x_{max} - x_{min}} \dots \text{wenn die Skala nicht mit 0 beginnt.}$$

P_i	Itemschwierigkeit
0,4 bis 0,6	niedrig
0,2 bis 0,4	mittelmäßig
0,6 bis 0,8	
< 0,2	hoch
> 0,8	

Im Sinne einer besseren Auswertbarkeit wird eine niedrige bis mittelmäßige Itemschwierigkeit angestrebt.

Überprüfen der Eindimensionalität einer Skala - Reliabilitätsanalyse

Berechnen der Itemschwierigkeit

Item	n	Σx_i	x_{\max}		P_i	Eignung
<i>Item d Inhaltliche Vollständigkeit</i>	73	334	5	4,58	0,895	<i>schlecht</i>
<i>Item a korrekte Terminologie</i>	75	316	5	4,21	0,802	<i>(knapp) mittelmäßig</i>
<i>Item e korrekte Grammatik</i>	75	311	5	4,15	0,787	<i>mittelmäßig</i>
<i>Item b flüssige Wiedergabe</i>	75	306	5	4,08	0,770	<i>mittelmäßig</i>
<i>Item c Native Aussprache</i>	75	184	5	2,45	0,363	<i>(knapp) gut</i>

Hohe Itemschwierigkeit = geringe Variation der Antworten
/Messwerte über die (Antwort-)Skala.

Siehe *Item d „Inhaltliche Vollständigkeit“*:

82% der Stichprobe befinden sich in einer Kategorie.

Überprüfen der Eindimensionalität einer Skala - Reliabilitätsanalyse

Reliabilitätsstatistiken		
Cronbachs Alpha	Cronbachs Alpha für standardisierte Items	Anzahl der Items
,806	,830	5

Cronbachs Alpha	Eignung der Items zur Indexbildung
> 0,9	exzellent
> 0,8	gut
> 0,7	akzeptabel
> 0,6	fragwürdig
> 0,5	schlecht
≤ 0,5	inakzeptabel

vgl. Blanz M.: Forschungsmethoden und Statistik für die Soziale Arbeit, 2015

Inter-Item-Korrelationsmatrix

	a	b	c	d	e
a korrekte Terminologie	1	,810	,429	,522	,549
b flüssige Wiedergabe	,810	1	,339	,551	,624
c native Aussprache	,429	,339	1	,221	,295
d Inhaltliche Vollständigkeit	,522	,551	,221	1	,598
e korrekte Grammatik	,549	,624	,295	,598	1

Überprüfen der Eindimensionalität einer Skala - Reliabilitätsanalyse

Mittelwert des Index, wenn das Item weggelassen wird

„Trennschärfe“ – Korrelation des Items mit dem Index aus den jeweils anderen 4 Items

Cronbachs Alpha, wenn das Item weggelassen wird

Item-Skala-Statistiken

	Skalen- mittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala- Korrelation	Quadrierte multiple Korrelation	Cronbach's Alpha, wenn Item weggelassen
a korrekte Terminologie	15,46	11,614	,752	,690	,726
b flüssige Wiedergabe	15,57	11,379	,741	,707	,726
c native Aussprache	17,24	11,752	,381	,193	,862
d Inhaltliche Vollständigkeit	15,10	12,671	,575	,420	,776
e korrekte Grammatik	15,49	11,819	,642	,488	,755

Überprüfen der Eindimensionalität einer Skala - Reliabilitätsanalyse

Formel zur Indexberechnung

$$\text{Index} = (a + b + c + d + e - 5) / (25 - 5) \cdot 100$$

The diagram illustrates the formula for calculating an index score. The formula is $\text{Index} = (a + b + c + d + e - 5) / (25 - 5) \cdot 100$. A purple bracket underlines the sum of items $(a + b + c + d + e)$. Three purple arrows point from the explanatory text boxes to the formula: one to the sum of items, one to the subtraction of 5, and one to the multiplication by 100.

Die Summation der 5 Items ergibt bei einer Skala von 1-5 einen Wertebereich von 5-25 für den Index.

Der kleinste Wert (min=5) wird abgezogen, das verankert den Index im Nullpunkt. (Wenn die Skala im Original bereits mit 0 beginnt, dann ist dieser Schritt nicht notwendig.)

Dann wird dividiert durch (max - min), also größter Wert 25 minus kleinster Wert 5 (in unserem Beispiel also 20), jetzt liegen alle Werte zwischen 0-1.

Die Multiplikation mit 100 transformiert die Skala auf 0 bis 100.

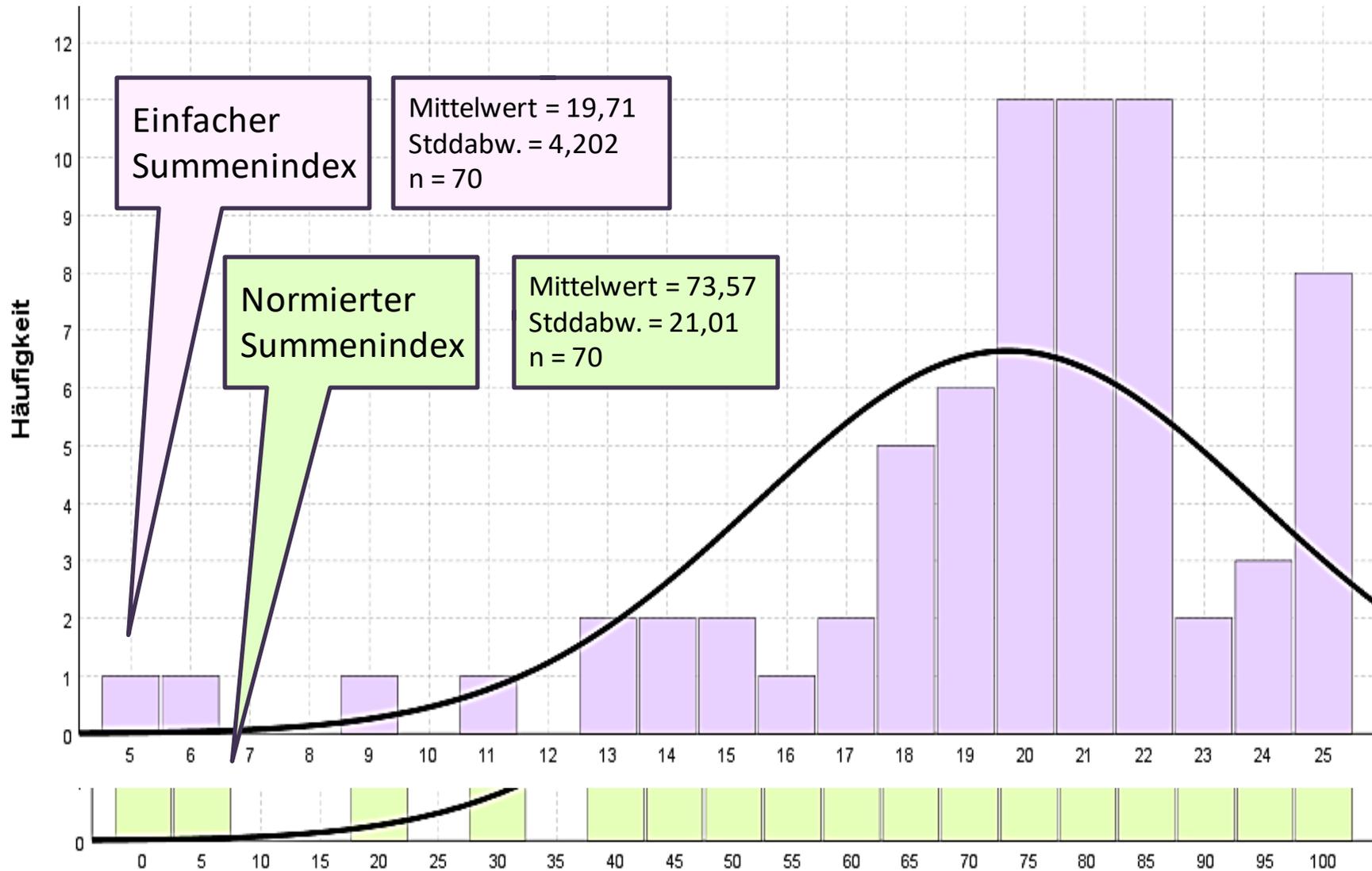
Index: Zufriedenheit mit der Dolmetschung

			Anz.	%	Gültige %	Kum. %
Gültig	5	0	1	1,2	1,4	1,4
	6	5	1	1,2	1,4	2,9
	9	20	1	1,2	1,4	4,3
	11	30	1	1,2	1,4	5,7
	13	40	2	2,5	2,9	8,6
	14	45	2	2,5	2,9	11,4
	15	50	2	2,5	2,9	14,3
	16	55	1	1,2	1,4	15,7
	17	60	2	2,5	2,9	18,6
	18	65	5	6,2	7,1	25,7
	19	70	6	7,4	8,6	34,3
	20	75	11	13,6	15,7	50,0
	21	80	11	13,6	15,7	65,7
	22	85	11	13,6	15,7	81,4
	23	90	2	2,5	2,9	84,3
	24	95	3	3,7	4,3	88,6
	25	100	8	9,9	11,4	100,0
	Gesamt		70	86,4	100,0	
Fehlend	System		11	13,6		
Gesamt			81	100,0		

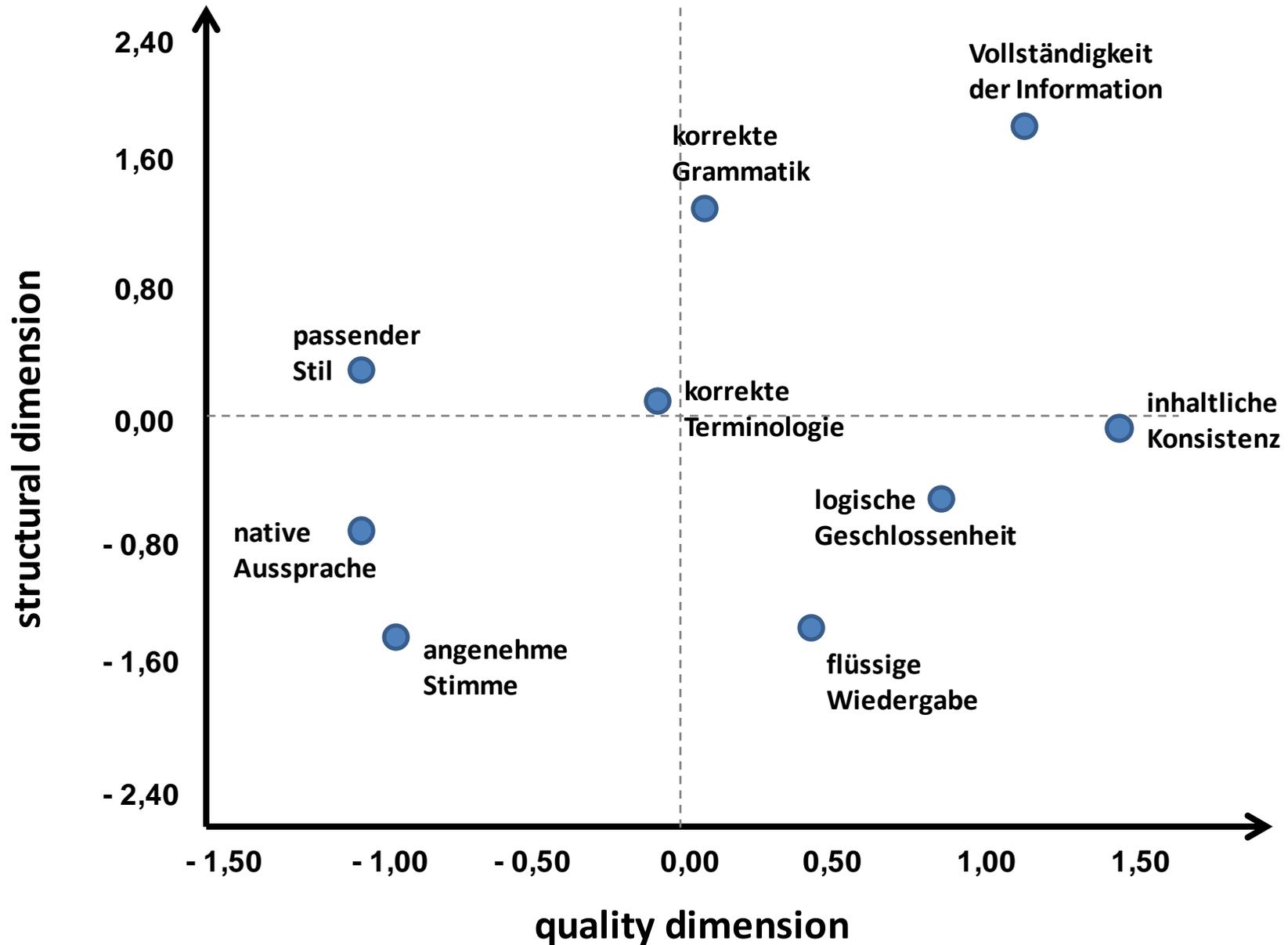
Skala des
einfachen
Summen-Index
5 - 25

Skala des
normierten
Summen-Index
0 - 100

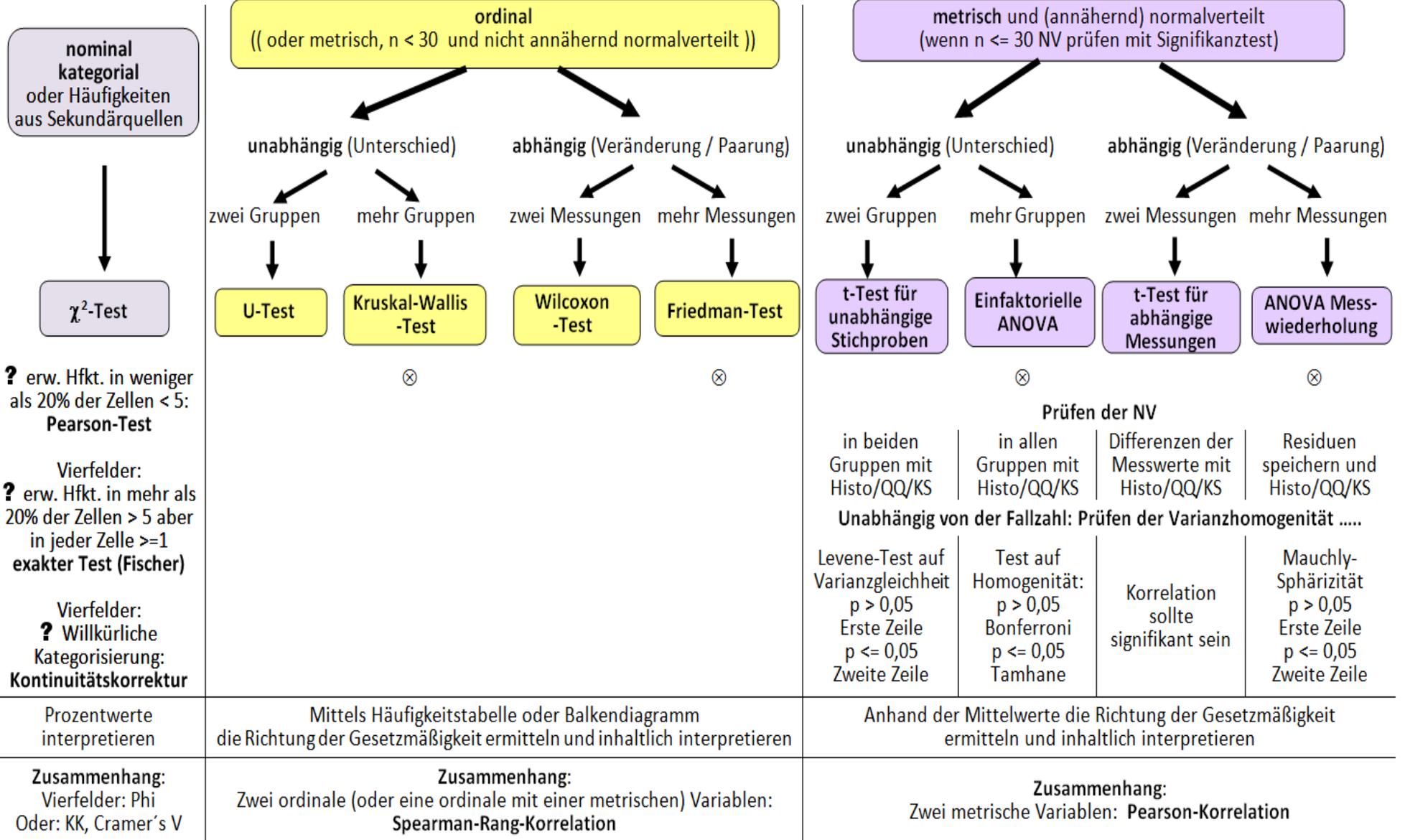
Zufriedenheit der Qualität von Dolmetschungen



Simultanübersetzung - Diskriminanzanalyse



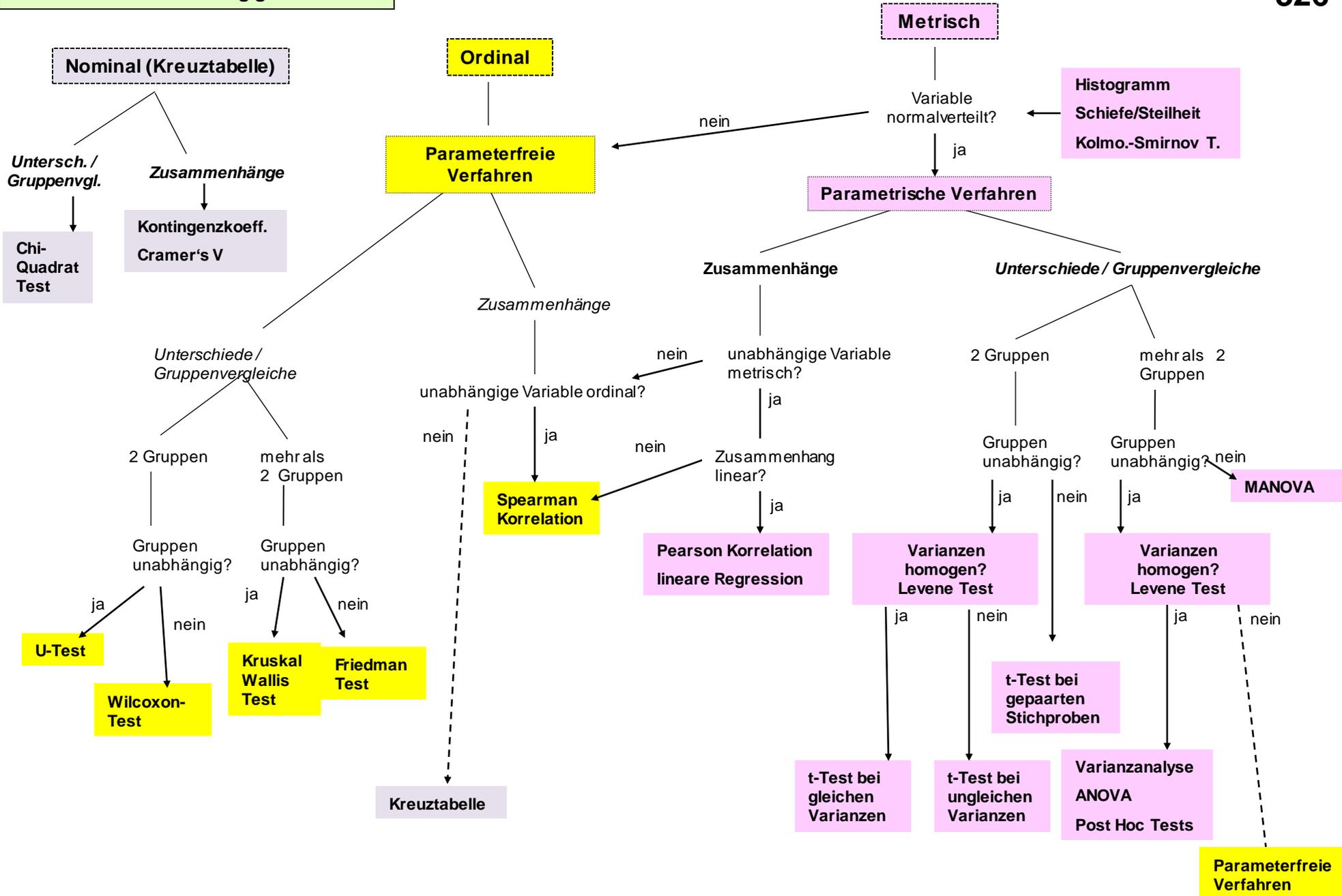
Entscheidungsbaum für statistische Verfahren



? erw. Hfkt. in weniger als 20% der Zellen < 5:
Pearson-Test

Vierfelder:
? erw. Hfkt. in mehr als 20% der Zellen > 5 aber in jeder Zelle ≥ 1
exakter Test (Fischer)

Vierfelder:
? Willkürliche Kategorisierung:
Kontinuitätskorrektur



Prüfung



Viel Erfolg bei
der Prüfung !!

Take-Home-Exam:
Formulieren einer sprachwissenschaftlichen Fragestellung
– ähnlich optionale Hausarbeit

Vorbereitung für die Prüfung:
Erarbeiten Sie die Prüfungsfragen mithilfe des Buches und der Folien
<https://homepage.univie.ac.at/isabella.hager/sprachwissenschaft/master>