

# Grundlagen der Diagnostik

## Sitzung 8

### **Einzelfalldiagnostik I**



We are happy to share our materials openly:

The content of these Open Educational Resources by Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München is licensed under CC BY-SA 4.0. The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

Sitzung	Datum	Thema	Themenblock
1	16.04.26	Einführung I	Definitionen; diagnostischer Prozess; gesetzlicher Rahmen; diagnostische Entscheidungen; Gütekriterien
2	23.04.26	Einführung II	
3	30.04.26	Einführung II fertig	
4	07.05.26	Verhaltensbeobachtung	Verhaltensbeobachtung als diagnostisches Verfahren
/	14.05.26	entfällt wegen Feiertag	
5	21.05.26	Beobachterübereinstimmung I	Maße zur Bestimmung der Übereinstimmung zwischen verschiedenen Beobachtern/Ratern
6	28.05.26	Beobachterübereinstimmung II	
/	04.06.26	<i>entfällt wegen Feiertag</i>	
7	11.06.26	Interviews	Interviews als diagnostisches Verfahren
8	18.06.26	Urteile und Fehler	Diagnostische Urteilsbildung und Güte von Urteilen
9	25.06.26	Einzelfalldiagnostik I	Methoden der Einzelfalldiagnostik aus der frequentistischen

- ➔ In der heutigen Vorlesung beschäftigen wir uns mit wichtigen Überlegungen in der Einzelfalldiagnostik, d.h. bei der Beurteilung einzelner konkreter Personen (z.B. Klientinnen, Bewerberinnen).

# 1. Wichtig Überlegung zur Messgenauigkeit

## Die Güte der Einzelfalldiagnostik hängt von der Güte der gesammelten Informationen ab!

- Die gesammelten Informationen sind die Grundlage für die Beantwortung einer diagnostischen Fragestellung und die damit einhergehenden Entscheidungen. Sie ermöglichen Aussagen über die Ausprägung einer Eigenschaft bei einzelnen Personen.
- Hierbei gilt es zu bedenken, dass die Messinstrumente der psychologischen Diagnostik nicht perfekt sind (siehe Gütekriterien!). Dies sollte bei der Auswahl der Instrumente, aber auch bei der Auswertung und Interpretation der Ergebnisse berücksichtigt werden.
  - ➔ **Je objektiver, reliabler, und valider das Messinstrument, desto vertrauenswürdiger das Ergebnis!**

## Praktische Relevanz der Reliabilität

Die Messgenauigkeit eines Tests muss bei der Beurteilung für den Einzelfall unbedingt berücksichtigt werden!

→ **Konfidenzintervalle drücken die Unsicherheit der Punktschätzung aus**

**Der in der Praxis leider häufig zu beobachtende Verzicht auf Konfidenzintervalle stellt einen groben Fehler dar und ist nicht entschuldbar!**

→ vgl. Gutachten-Standards der Föderation Deutscher Psychologenverbände

## 2. Arten der Testauswertung

## Testauswertung

Es gibt zwei verschiedene Möglichkeiten das Testergebnis einer Person auszuwerten:

- **Kriteriumsorientiert**: Das Testergebnis wird mit einem **feststehenden Kriterium** verglichen, z.B. muss für eine positive Entscheidung ein bestimmter absoluter Wert erreicht werden
- **Normorientiert**: Das Testergebnis wird **in Relation zu einer (sinnvollen) Bezugsgruppe** interpretiert, z.B. muss für eine positive Entscheidung ein überdurchschnittliches Ergebnis im Vgl. zu Gleichaltrigen vorliegen

### Kriteriumsorientierte Testauswertung

Beispiel: In der theoretischen Führerscheinprüfung dürfen für das Bestehen nicht mehr als 10 Fehlerpunkte gemacht werden

Hierfür benötigt man also...

- den Testwert i.d.R. als Rohwert
- ein Cut-Off-Kriterium

In der psychologischen Diagnostik sollte sich das Cut-Off-Kriterium im besten Fall auf das Konfidenzintervall (nicht den Punktschätzer) beziehen (z.B. „untere Grenze des 95%-KI > 100“)

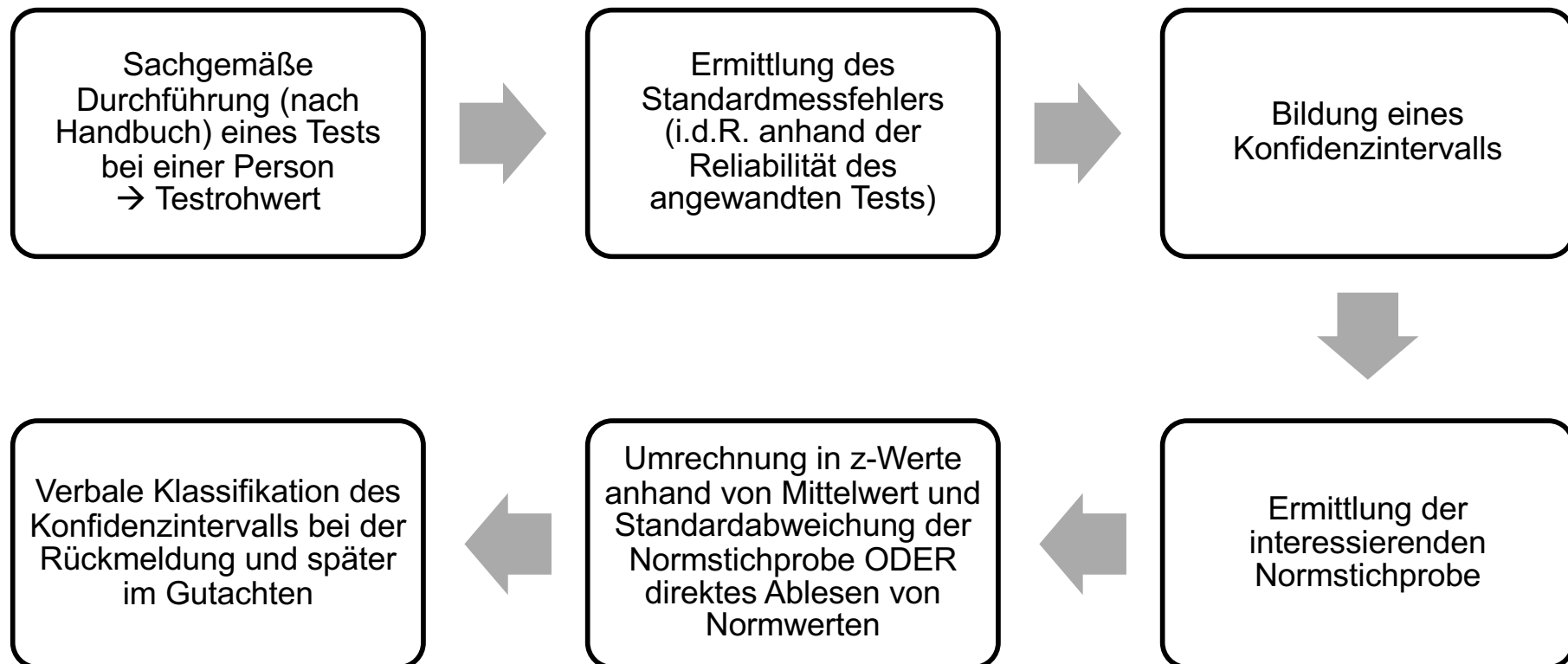
### Normorientierte Testauswertung

Beispiel: Nur Kinder, die im Vergleich zu Gleichaltrigen überdurchschnittliche Werte im Intelligenztest aufweisen, können eine Klasse überspringen

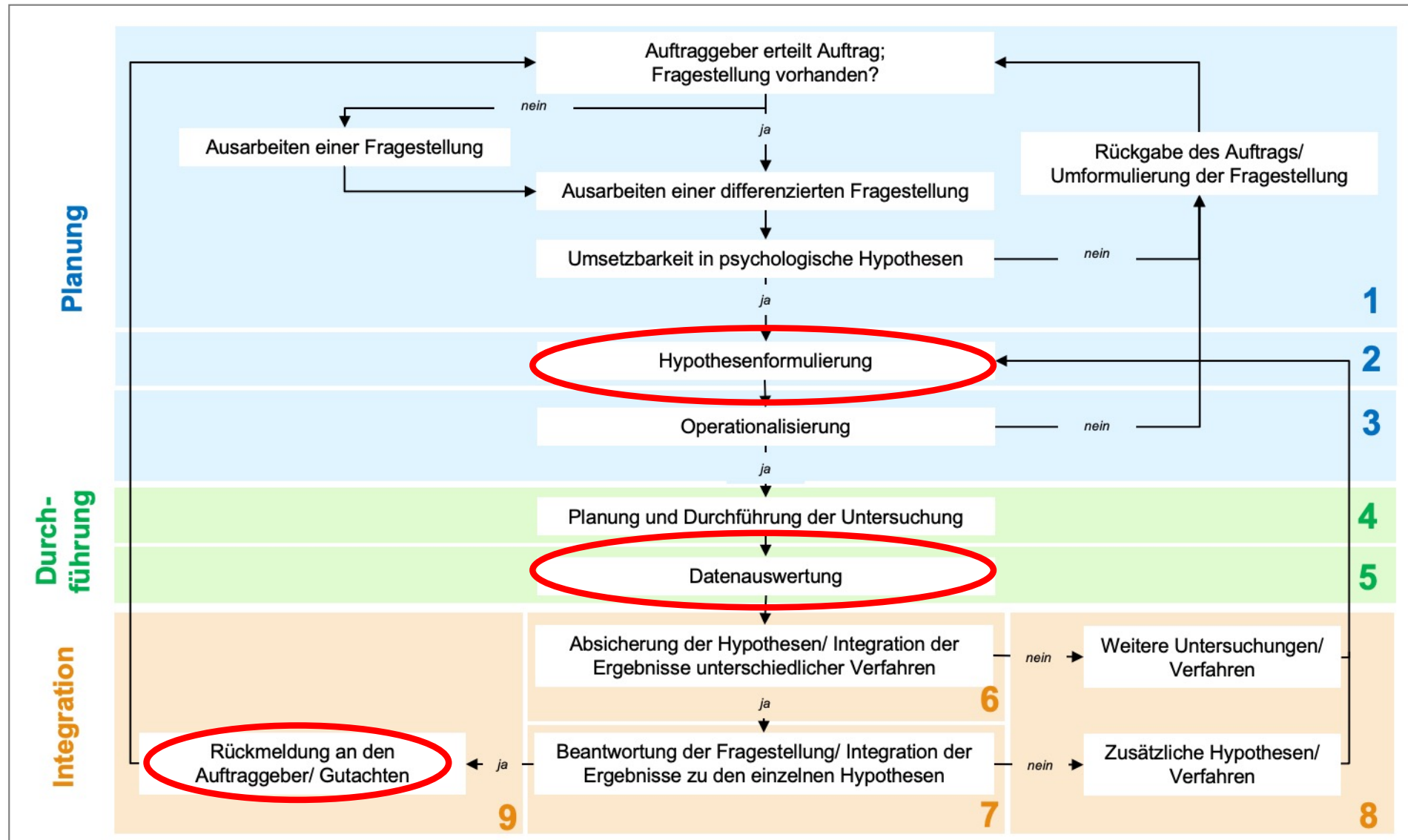
Hierfür benötigt man also...

- den Testwert i.d.R. als Normwert (entweder über Umrechnung aus Rohwert oder direkt aus Normwerttabelle)
- das Konfidenzintervall für den Normwert
- detaillierte Informationen zur passenden Bezugsgruppe („Normstichprobe“)

## Schritte der **normorientierten** Einzelfalldiagnostik (am Bsp. eines Tests)

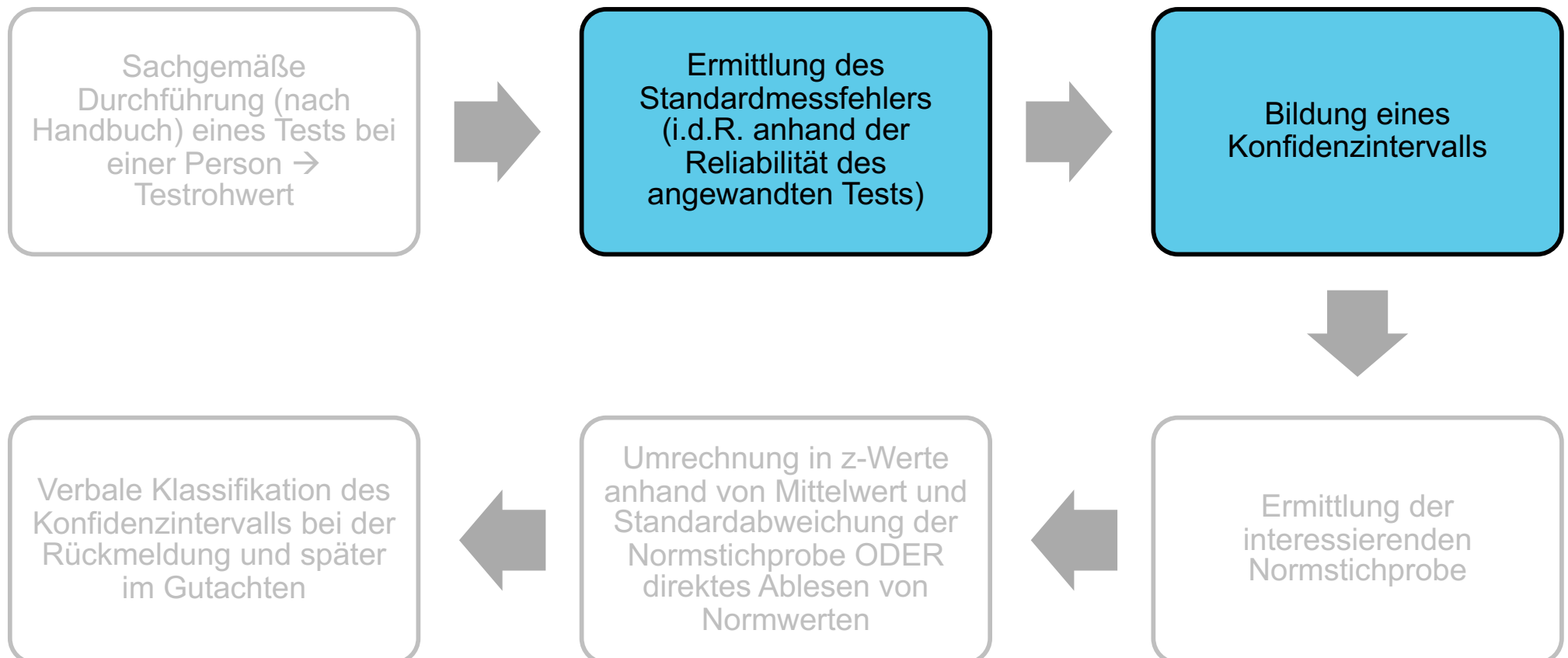


Hinweis: In der Praxis werden die Rohwerte teilweise zuerst in Normwerte überführt, bevor das Konfidenzintervall gebildet wird. Die oben beschriebene Reihenfolge ist allerdings empfehlenswerte!



# 3. Berechnung von frequentistischen Konfidenzintervallen

## Schritte der **normorientierten** Einzelfalldiagnostik (am Bsp. eines Tests)



## Konfidenzintervall für Summen- und Mittelwerte mit den **in der Praxis** vorhandenen Informationen (siehe VL13-14 der Testtheorie!)

### 1. Approximativer Standardmessfehler: $SE(\hat{\theta}_{Person}) = \sigma_x * \sqrt{(1 - Rel(X))}$

- $\sigma_x$  = Standardabweichung des Summen- / Mittelwerts
  - Wenn man das KI für einen Rohwert berechnet, dann ist  $\sigma_x$  in der Einheit der Rohwerte
  - Wenn man das KI für einen Normwert berechnet, dann ist  $\sigma_x$  in der Einheit der Normwerte (z.B. 15 bei IQ-Werten oder 1 bei z-Werten)
- Die Reliabilität selbst wird nur geschätzt und die Qualität der Schätzung hängt von der Passung der verwendeten Methode (z.B. Cronbachs Alpha) und der Stichprobe (Größe, Repräsentativität) ab!

## Konfidenzintervall für Summen- und Mittelwerte mit den **in der Praxis** vorhandenen Informationen (siehe VL13-14 der Testtheorie!)

### 2. Zweiseitige KI-Berechnung mit kritischem z-Wert:

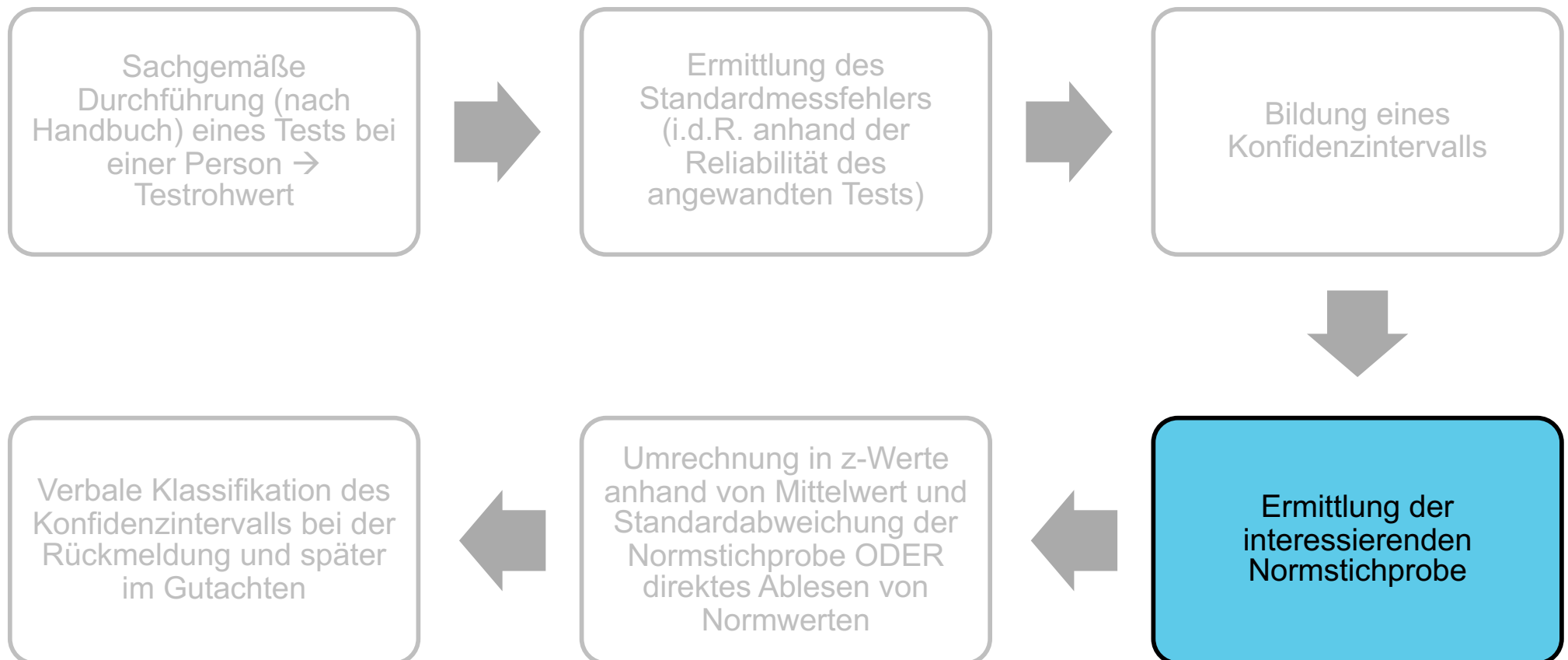
$$(1 - \alpha)\% \text{-KI} = x \pm z_{1-\alpha/2} * SE(\hat{\theta}_{Person}) = x \pm z_{1-\alpha/2} * \underbrace{\sigma_x * \sqrt{(1 - Rel(X))}}_{\text{Approximativer Standardmessfehler}}$$

- 95%-KI =  $x \pm 1.96 * SE(\hat{\theta}_{Person})$
- 90%-KI =  $x \pm 1.64 * SE(\hat{\theta}_{Person})$
- 80%-KI =  $x \pm 1.28 * SE(\hat{\theta}_{Person})$

→ Die Breite des Konfidenzinterfalls für Testwerte in der Einzelfalldiagnostik hängt sowohl von der Reliabilität des verwendeten, als auch vom gewählten Konfidenzniveau ab!

# 4. Beurteilung von Normstichproben

## Schritte der **normorientierten** Einzelfalldiagnostik (am Bsp. eines Tests)



## Welches ist die „interessierende“ Normstichprobe?

Welche Bezugsgruppe für meinen diagnostischen Fall interessant ist, hängt vom jeweiligen Kontext ab:

Es gibt Situationen, da möchte ich...

- **...eine Normstichprobe, die meiner getesteten Person möglichst ähnlich ist**, z.B. wenn ich die motorischen Fähigkeiten von Kindern untersuche, um Entwicklungsverzögerungen festzustellen  
→ hier sind sowohl geschlechts- als auch altersspezifische Unterschiede zu erwarten und wichtig zu berücksichtigen
- **...eine (bevölkerungsrepräsentative) Gesamtnormstichprobe und mir ist egal ob es Untergruppen mit Unterschieden auf dem gemessenen Merkmal gibt**, z.B. wenn ich in meinem Unternehmen jemanden mit hoher Konzentrationsfähigkeit einstellen möchte  
→ hier sind womöglich altersabhängige Unterschiede zu erwarten, sie spielen aber keine Rolle bei der Beurteilung

## Relevante Eigenschaften von Normstichproben

### Wichtigste Aspekte:

- Größe
- Aktualität
- Repräsentativität

### Weitere Aspekte:

- Es sind Teilnormstichproben vorhanden, die empirisch oder praktisch begründet sind (z.B. für verschiedene Altersgruppen bei kognitiven Leistungen) → Achtung: Fragen Sie sich, ob eine Unterteilung gerechtfertigt ist!
- Idealerweise gibt es für jede Teilnormstichprobe eine eigene Reliabilitätsschätzung – diese wird für die Berechnung des Standardmessfehlers benötigt

## Größe der Normstichprobe

- **Grund für Relevanz:** Je größer die Normstichprobe, desto präziser sind die Schätzungen der unbekanntes Populationsgrößen  $\mu$  (Mittelwert) und  $\sigma$  (Standardabweichung), die für die Berechnung von Normwerten herangezogen werden (siehe Abschnitt 5)
- **Empfehlung zur Mindestgröße:**
  - Faustregel: Die Stichprobe sollte mindestens 200-400 Personen umfassen!
  - Vorsicht: Für eine generelle Bewertung von Normen sind hier die Teilnormstichproben gemeint, nicht die Gesamtnormstichprobe!
  - Für den konkreten einzelfalldiagnostischen Fall ist die Größe der jeweils interessierenden Normstichprobe relevant!

## Aktualität der Normstichprobe

- **Grund für Relevanz:** Veränderungen auf dem Merkmal in der Population sind möglich
- **Empfehlung zum Maximalalter:** Der DIN-Norm 33430 zufolge, soll mindestens alle acht Jahre überprüft werden, ob die Normen noch aktuell sind (Kersting, 2007)

## Repräsentativität der Normstichprobe

Eine Stichprobe ist repräsentativ, wenn die Stichprobe hinsichtlich ihrer Zusammensetzung die Zielpopulation möglichst genau abbildet. Ob eine Stichprobe repräsentativ ist, wird durch die Art der Ziehung sichergestellt.

- **Grund für Relevanz:** Die Normstichprobe soll Aussagen in Bezug auf eine Population ermöglichen
- **Empfehlung zum optimalen Vorgehen:**
  1. Definition der Zielpopulation
  2. Ziehung einer Zufallsstichprobe

## Repräsentativität der Normstichprobe

- **Aber:** Die Ziehung einer Zufallsstichprobe ist in der Praxis höchst selten umsetzbar.
  - **Empfehlung zur Beurteilung:** Überprüfung von...
    - Angaben zur Rekrutierung (Anwerbung, Rücklaufquoten, Belohnung)  
→ Mögliche Verzerrungen bedenken!
    - Stichprobenzusammensetzung, z.B. im Vergleich zu Zensusdaten
    - Angaben zur Testsituation (z.B. Einzel- oder Gruppentestung, Bewerbungssituation, Teil einer Testbatterie, fand Ergebnisrückmeldung statt)  
→ Mögliche Einflüsse auf Testergebnisse bedenken!
- Fehlende Angaben sind problematisch, da überhaupt keine Abschätzung der Repräsentativität möglich ist!

# Stichprobenzusammensetzung: Beispiel (NEO-PI-R)

**Tabelle 15:** Soziodemographische Merkmale der Probanden der nicht-klinischen Stichprobe und der klinischen Stichprobe. Zum Vergleich sind Statistiken angegeben, die auf den Angaben des Statistischen Jahrbuchs des Bundesamts für Statistik für das Jahr 2001 beruhen.

Soziodemographische Merkmale	NKS %	SB Statistik %	KS %
Geschlecht			
männlich	35.99	48.3	50.2
Weiblich	64.01	51.7	49.8
Alter			
16-18	2.41	2.69	
18-20	7.86	2.76	
20-22	16.35	2.66	
22-24	13.92	2.66	
24-26	10.82	2.64	
26-28	7.70	2.87	
28-30	5.60	3.43	
30-35	9.66	10.16	
35-40	6.17	10.45	
40-45	5.13	9.09	
45-50	4.86	8.28	
50-55	3.57	6.75	
55-60	2.89	7.94	
60-65	1.42	8.09	
65-70	.93	5.87	
70-80	.57	9.37	
>= 80	.13	4.29	
Minimum - Maximum	16-91		16-70
Mittelwert	29.92	47.39	36.80
Standardabweichung	12.08	18.47	12.88
Schulabschluss			
Vor letzter Hauptschulklasse	.99	2.6	12.8
Mit letzter Hauptschulklasse	6.99	46.9	24.4
Realschule ohne Abschluss	1.34		5.0
Realschule mit Abschluss	16.78	26.6	20.6
Höhere Schule ohne Abitur	6.91		11.1
Abitur ohne Studium	13.47		6.1
Abitur und anschl. Studium	40.21	19.3	9.4
Abitur u. abgeschl. Studium	13.20		10.6
ohne Angabe: <i>N</i> = 1 966			<i>N</i> = 99
< Abitur	33.02	80.7	73.89
>= Abitur	66.98	19.3	26.11
ohne Angabe: <i>N</i> = 1 966			<i>N</i> = 99
Land (nach Wohnort)			
Westliche Bundesländer	68.61		99.6
Östliche Bundesländer	25.39		.4
Österreich	4.96		-
Schweiz	1.04		-
ohne Angabe: <i>N</i> = 72			<i>N</i> = 0
BRD Probanden (nach Wohnort)			
Westliche Bundesländer	72.99	81.48	99.6
Östliche Bundesländer	27.01	18.52	.4
Deutsche Bundesländer (Wohnort)			
Baden-Württemberg	7.0	12.75	
Bayern	5.6	14.79	
Berlin	2.3	4.12	

Anpassung



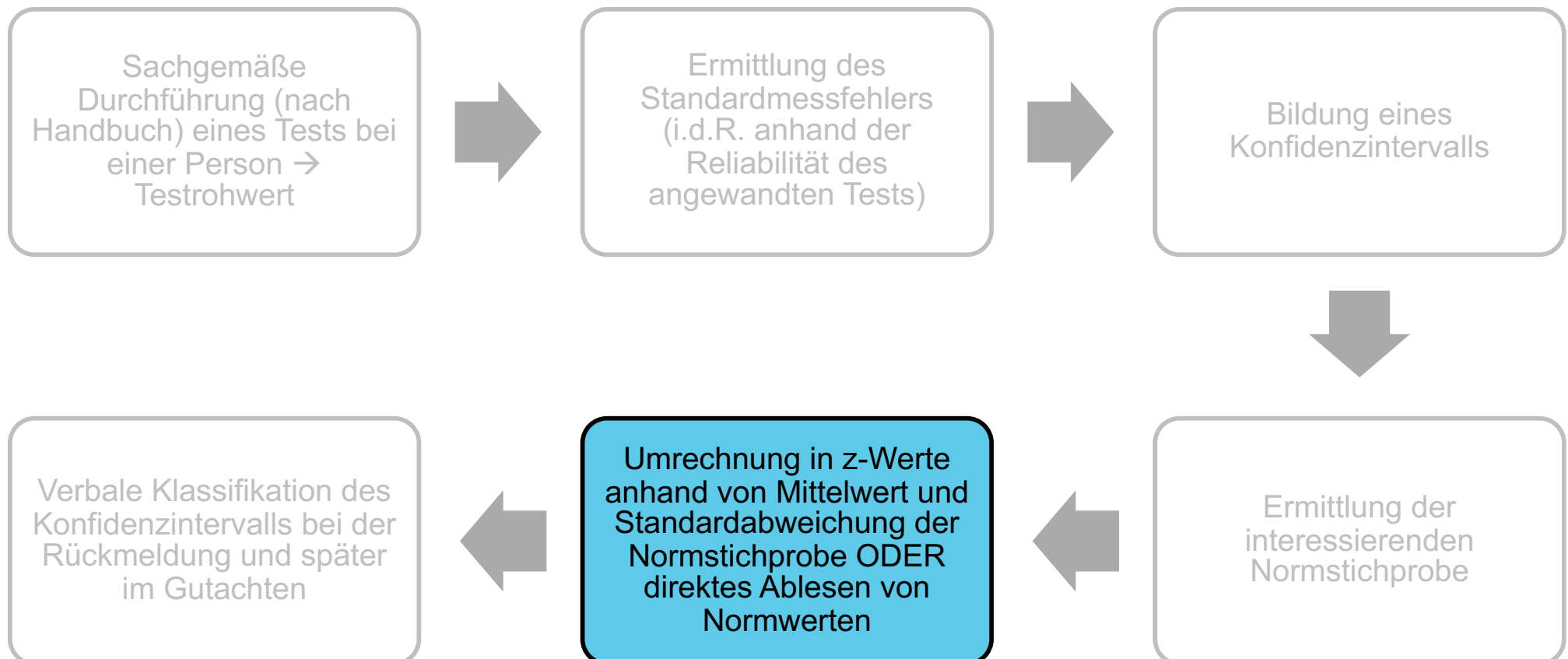
**Tabelle 37:** Soziodemographische Merkmale der Probanden repräsentativen Stichprobe (*N* = 871). Zum Vergleich sind Statistiken angegeben, die auf den Angaben des Statistischen Jahrbuchs des Bundesamts für Statistik für das Jahr 2001 beruhen.

Soziodemographische Merkmale	Quotenstichprobe %	SB Statistik %
Geschlecht		
männlich	48.6	48.3
Weiblich	51.4	51.7
Alter (von ... bis unter ...)		
16-18	2.64	2.69
18-20	2.76	2.76
20-22	2.64	2.66
22-24	2.07	2.66
24-26	3.21	2.64
26-28	2.53	2.87
28-30	3.79	3.43
30-35	10.22	10.16
35-40	10.45	10.45
40-45	9.07	9.09
45-50	8.27	8.28
50-55	6.77	6.75
55-60	9.29	7.94
60-65	9.87	8.09
65-70	8.61	5.87
70-80	6.20	9.37
>= 80	1.61	4.29
Minimum - Maximum	16-91	
Mittelwert	46.15	47.39
Standardabweichung	16.89	18.47
Schulabschluss		
< Abitur	80.3	80.4
>= Abitur	19.7	19.6

**Anmerkung:** SB-Statistik = Werte basierend auf den Angaben des Statistischen Bundesamtes (Statistisches Jahrbuch 2001).

# 5. Berechnung von Normwerten und Prozenträngen

## Schritte der **normorientierten** Einzelfalldiagnostik (am Bsp. eines Tests)



## Bildung von z-Wert basierten Normwerten: Schritt 1

- Die Testrohwerte werden zunächst mittels linearer Transformation in z-standardisierte Werte überführt:

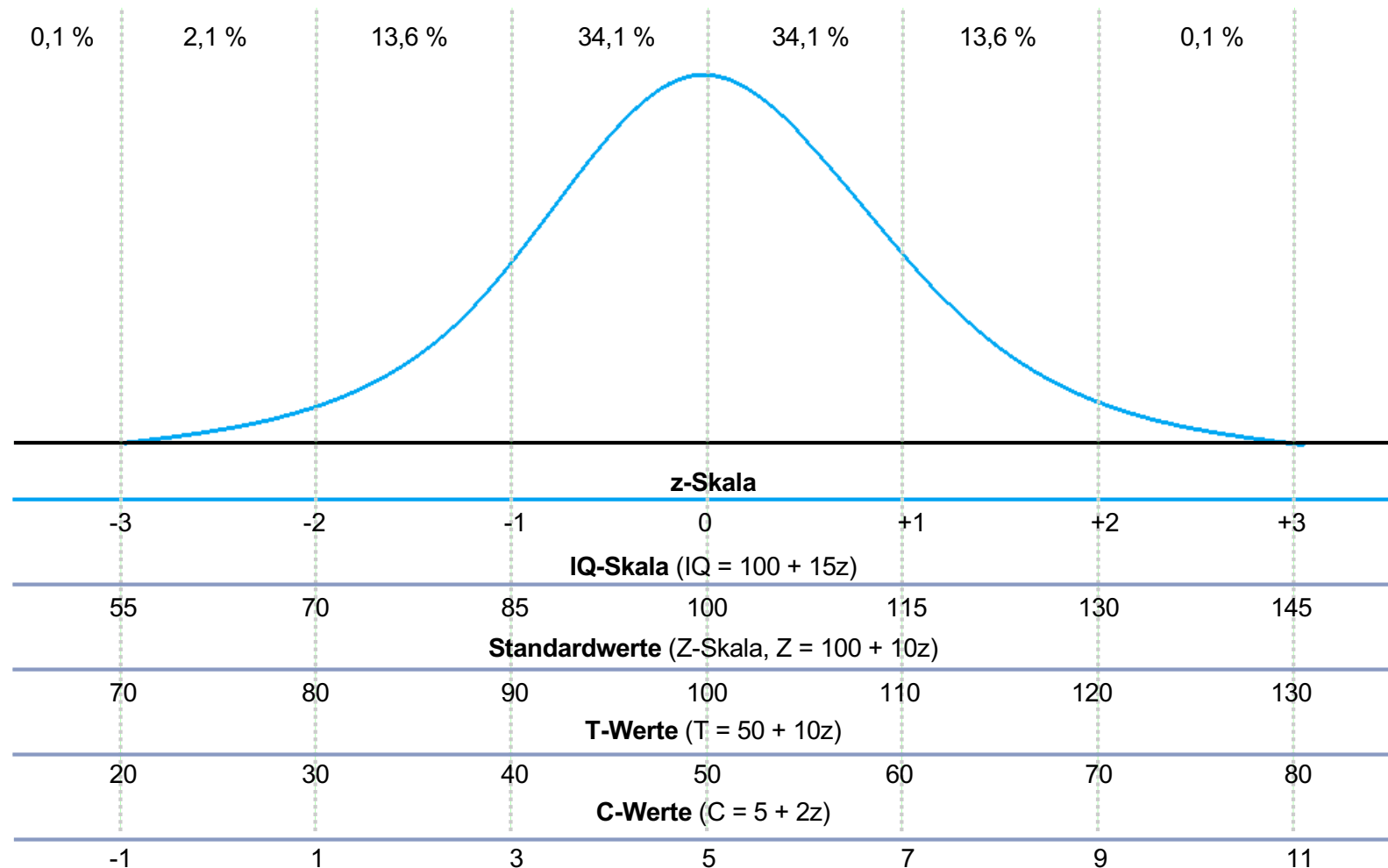
$$z_v = \frac{x_v - \bar{x}}{s}$$

- $z_v$  = z-standardisierter Wert der Person v
- $x_v$  = Testrohwert der Person v
- die Schätzwerte  $\bar{x}$  und  $s$  sind in der Einheit der Testrohwerte und stammen aus der interessierenden Normstichprobe

zur WH:

- ein z-Wert gibt in Standardabweichungen an, wie groß die Abweichung eines Testwerts vom Mittelwert ist
- z-Werte haben einen Mittelwert von 0 & eine Standardabweichung von 1

### Verschiedene Normwertskalen



## Bilden von z-Wert-basierten Normwerten: Schritt 2

- Die z-standardisierte Werte werden nun in Normwerte überführt

Generell: Normwert = MW der Normskala + (SD der Normskala \*  $z_v$ )

Wert	Berechnung
IQ-Werte	$IQ = 100 + (15 * z_v)$
Standardwerte	$SW = 100 + (10 * z_v)$
T-Werte	$T = 50 + (10 * z_v)$
C-Werte	$C = 5 + (2 * z_v)$

- Alle Normwerte lassen sich ineinander überführen! Der Informationsgehalt aller Normwerte ist identisch, sie befinden sich nur auf einer anderen Skala!**
- Vorsicht: Normwerte haben ohne Kenntnis der genutzten Bezugsgruppe nur eine eingeschränkte Bedeutung!

## Beispiel für eine Normtabelle (NEO-PI-R)

Männer mit hohem Bildungsgrad

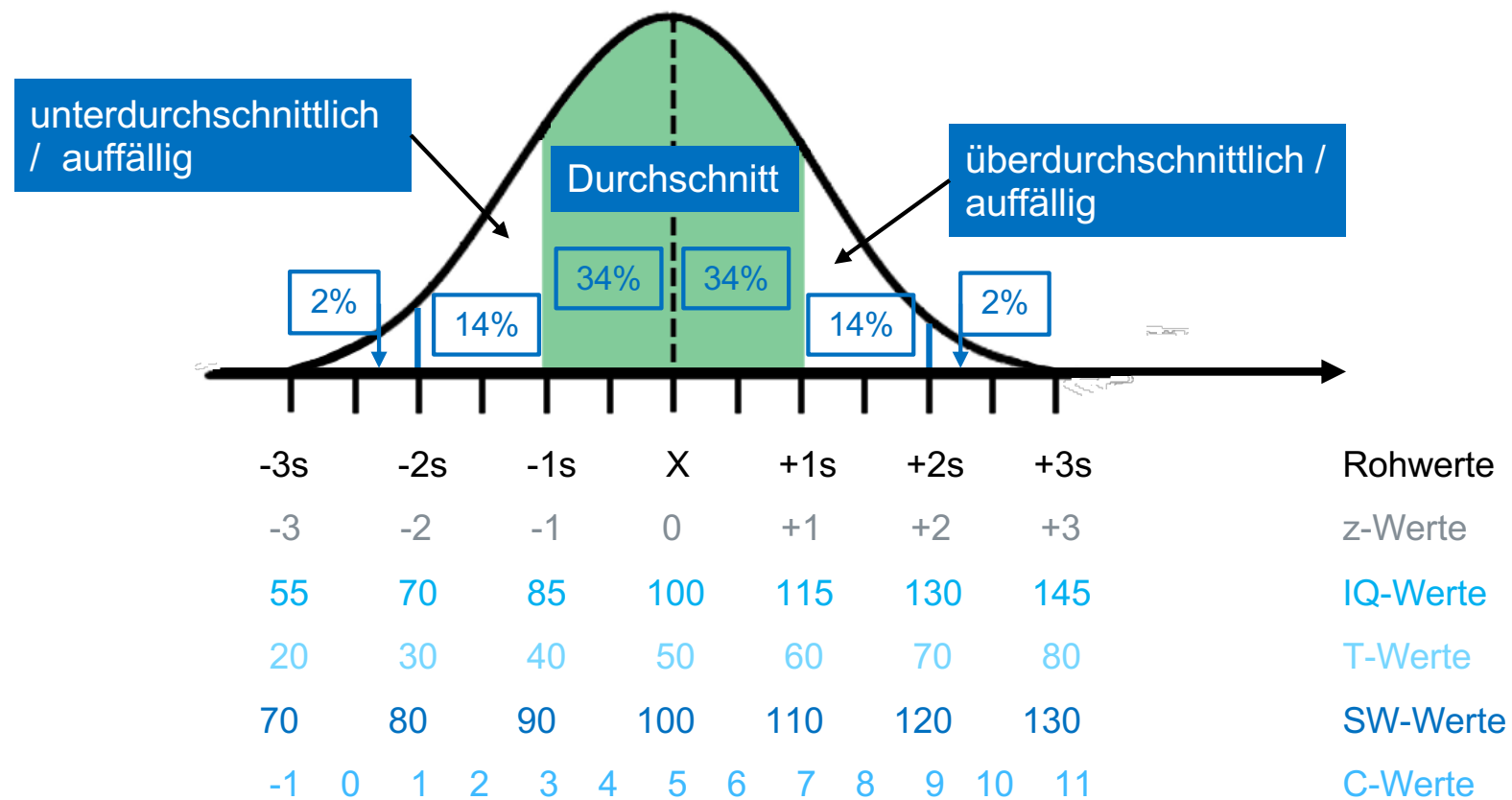
NEO-PI-R-Form S Normwerte

Tabelle A: Männer im Alter von 16–29 Jahren mit Hochschul- bzw. Fachhochschulreife (N=1709)

T	O	O1	O2	O3	O4	O5	O6	T	ST	PR	
80	177-192				30-32			80	9	99.9	
79	176						31-32	79	9	99.8	
78	172-175							78	9	99.7	
77	171	32	32	32	29	32		77	9	99.7	
76	169-170							76	9	99.5	
75	168							75	9	99.4	
74					28		30	74	9	99.2	
73	167		31					73	9	98.9	
72	164-166	31						72	9	98.6	
71	163			31	27	31	29	71	9	98.2	
70	161-162		30					70	9	97.7	
69	159-160				26			69	9	97.1	
68	157-158	30		30		30	28	68	9	96.4	
67	155-156		29					67	8	95.5	
66	154	29			25	29		66	8	94.5	
65	152-153		29				27	65	8	93.3	
64	150-151		28		24			64	8	91.9	
63	148-149	28		28		28	26	63	8	90.3	
62	147							62	7	88.5	
61	145-146	27	27	27	23	27		61	7	86.4	
60	143-144	26					25	60	7	84.1	
59	141-142		26		22	26		59	7	81.6	
58	139-140	25		26				58	7	78.8	
57	137-138		25		21	25	24	57	6	75.8	
56	135-136	24		25				56	6	72.6	
55	133-134		24		20	24		55	6	69.2	
54	131-132	23		24			23	54	6	65.5	
53	130		23		19	23		53	6	61.8	
52	128-129	22						52	5	57.9	
51	126-127		22	23	18	22	22	51	5	54.0	
50	124-125	21						50	5	50.0	
49	122-123		21				21	21	49	5	46.0
48	121	20		22	17				48	5	42.1
47	119-120		20			20			47	4	38.2
46	117-118	19	19	21	16		20		46	4	34.5
45	114-116	18				19			45	4	30.9
44	113		18	20	15				44	4	27.5
43	111-112	17	17			18	19		43	4	24.2
42	109-110	16		19	14				42	3	21.2
41	107-108		16			17	18		41	3	18.4
40	105-106	15	15	18	13	16			40	3	15.9
39	104		14						39	3	13.6
38	101-103	14		17		15	17		38	3	11.5
37	99-100		13		12				37	2	9.7
36	97-98	13	12	16		14	16		36	2	8.1
35	95-96		11		11	13			35	2	6.7
34	92-94	12		15					34	2	5.5
33	90-91	11	10		10	12	15		33	2	4.5
32	89			14		11			32	1	3.6
31	87-88	10	9	13	9				31	1	2.9
30	85-86		12		10	14			30	1	2.3
29	82-84	9	8		8				29	1	1.8
28	81		7	11		9	13		28	1	1.4
27	78-80	8			7				27	1	1.1
26	76-77		6	10			12		26	1	0.8
25	74-75					8	11		25	1	0.6
24	72-73	7	5	9	6		10		24	1	0.5
23	69-71						7		23	1	0.4
22	68	6	4		5	6	7-9		22	1	0.3
21	66-67	5		7-8	3-4				21	1	0.2
20	0-65	0-4	0-3	0-6	0-2	0-5	0-6		20	1	<0.1

Anmerkung: T=T-Werte, ST=Stanine Werte, PR=Prozenträge; O=Offenheit für Erfahrungen, O1=Offenheit für Phantasie, O2=Offenheit für Ästhetik, O3=Offenheit für Gefühle, O4=Offenheit für Handlungen, O5=Offenheit für Ideen, O6=Offenheit des Werte- und Normensystems.

**Bei Annahme einer Normalverteilung befinden sich 68% aller Werte im Bereich von  $\pm 1SD$  vom Mittelwert:**



## Skalenniveau von z-Wert basierten Normwerten

- z-Wert basierte Normwerte haben Intervallskalenniveau\* und sind standardisiert
- Konsequenz:
  - Der Abstand zweier Normwerte spiegelt den Abstand in Testwerten wider:  $z = 2.5$  zu  $z = 2$  ist der gleiche Testwertabstand wie  $z = 0.5$  zu  $z = 0$
  - Es sind Analysen möglich, die Intervallskalenniveau voraussetzen (z.B. Pearson-Korrelationen, t-Tests)
  - Standardisierung: Vergleich verschiedener Testergebnisse einer Person oder mehrerer Personen möglich, denn z-Werte sind unabhängig von der Original-Maßeinheit der Skalen

\* weil sie Intervallskalenniveau der Testrohwerte voraussetzen

## Ergänzender Kennwert: Prozentränge

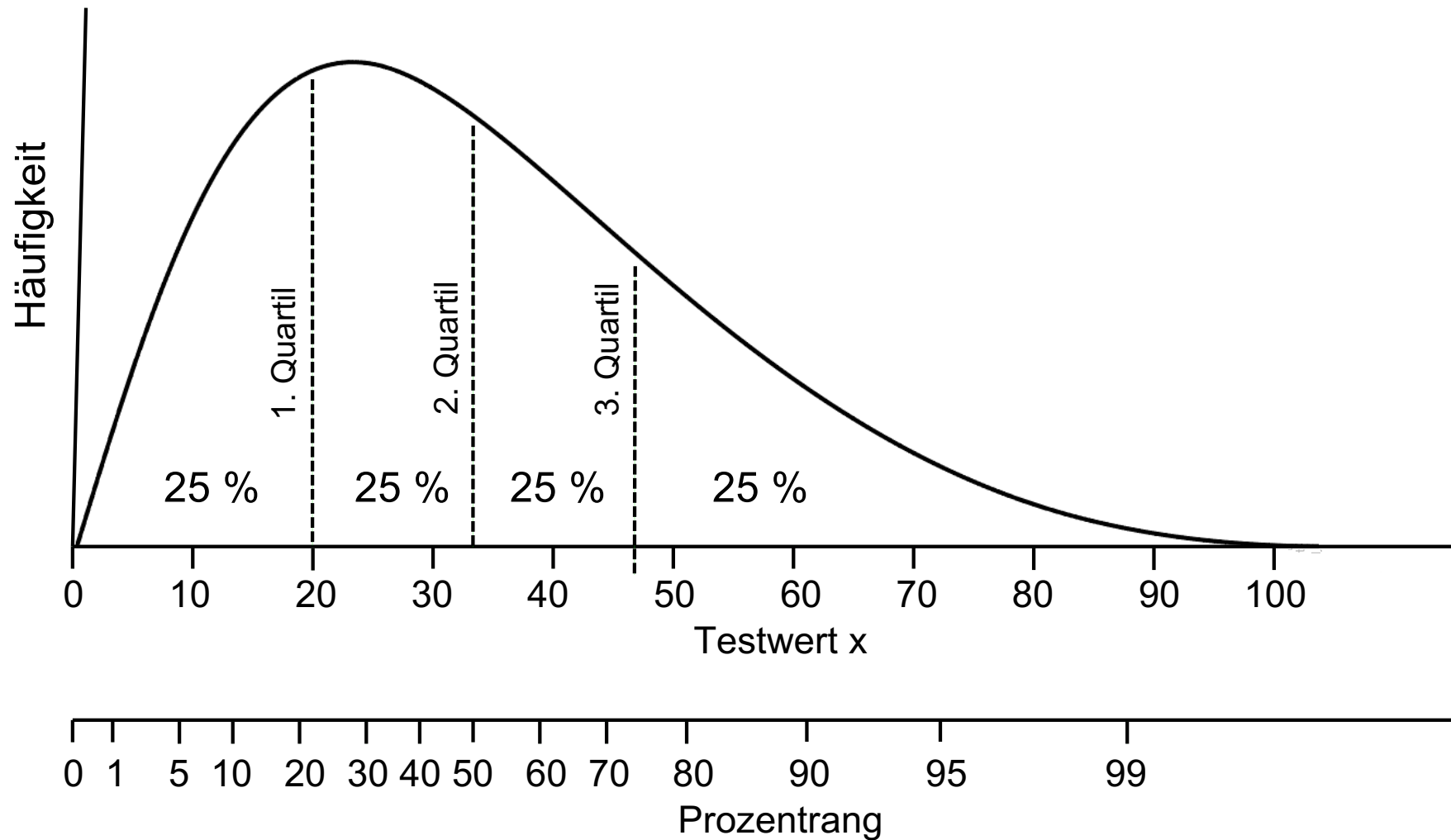
- Ein Prozentrang gibt an, wie viel Prozent der Normstichprobe einen Testwert erzielen, der niedriger oder maximal ebenso hoch ist, wie der Testwert der getesteten Person
- Bestimmung von Prozenträngen: Multiplikation der relativen kumulierten Häufigkeiten mit dem Faktor 100

Testwert	H	h	$h_{\text{kum}}$	(Geschätzter) Prozentrang
0	8	0,016	0,016	1,6
1	12	0,024	0,04	4
2	17	0,034	0,074	7,4
3	25	0,05	0,124	12,4
4	37	0,074	0,198	19,8
5	54	0,108	0,306	30,6
6	66	0,132	0,438	43,8
7	95	0,19	0,628	62,8
8	120	0,24	0,868	86,8
9	53	0,106	0,974	97,4
10	13	0,026	1	100

## Ergänzender Kennwert: Prozentränge

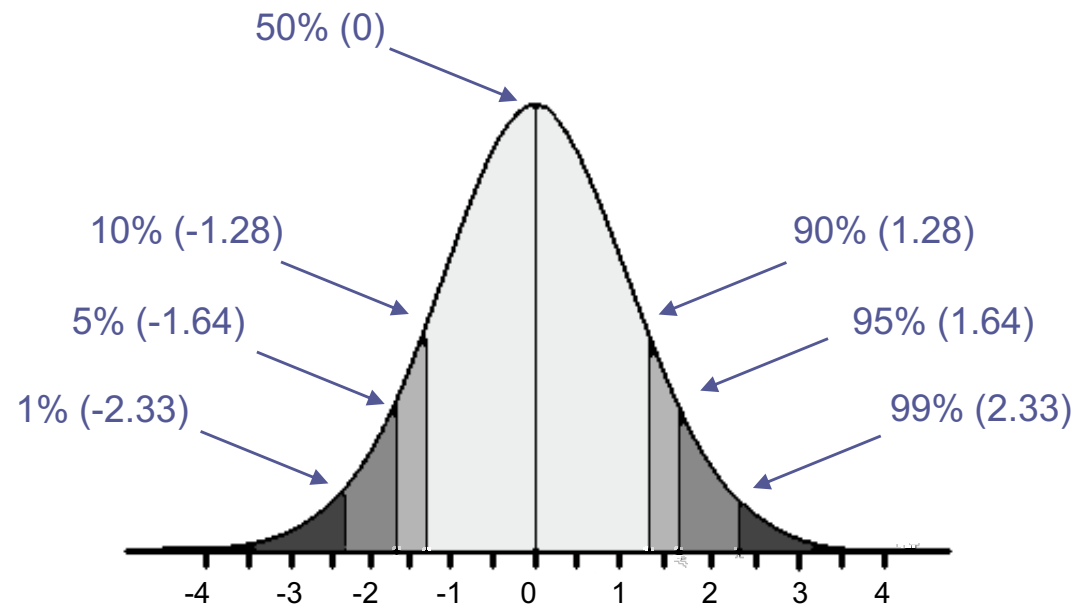
- Prozentränge eignen sich auch zur Beschreibung schief verteilter Werte
- Für Laien in der Regel gut verständlich, aber Prozentränge haben Ordinalskalenniveau und deshalb folgende Probleme:
  - Abstand zweier Prozentränge spiegelt nicht den Abstand in Testwerten wider → Vergleich zwischen Personen schwierig
  - Hohe Differenzierbarkeit in Bereichen, in denen es viele Werte gibt und niedrige Differenzierbarkeit in Bereichen, in denen es wenige Werte gibt

## Abstände zwischen Prozenträngen



### Abstände zwischen Prozenträngen

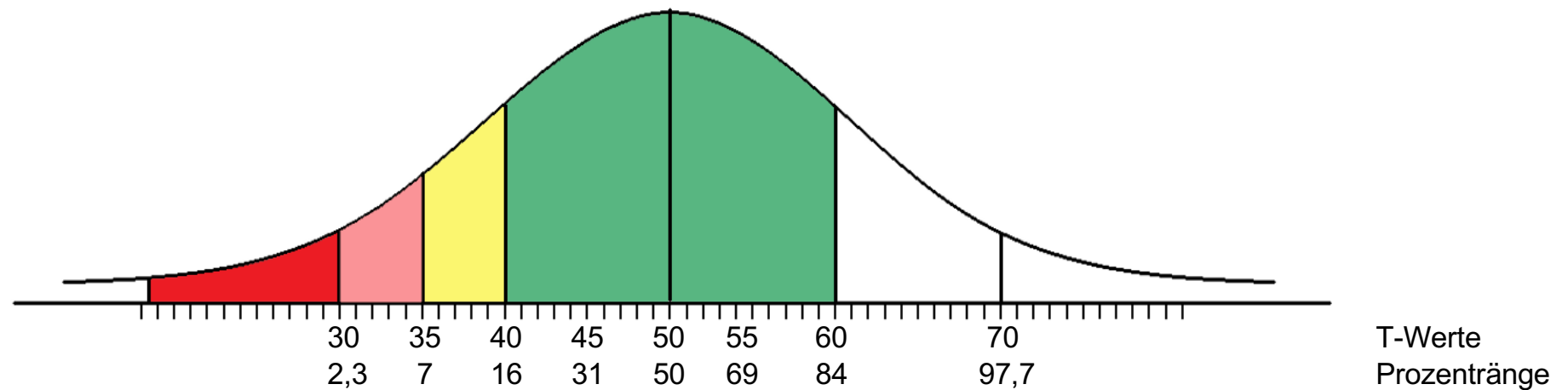
Bei Annahme einer Normalverteilung...



Berechnung der Prozentränge in R mit:  $\text{pnorm}(z_v, \text{mean}=0, \text{sd}=1) * 100$

Wo liegen hier die Prozentränge nah beieinander, wo weiter voneinander entfernt?

Nochmal zum Vergleich:

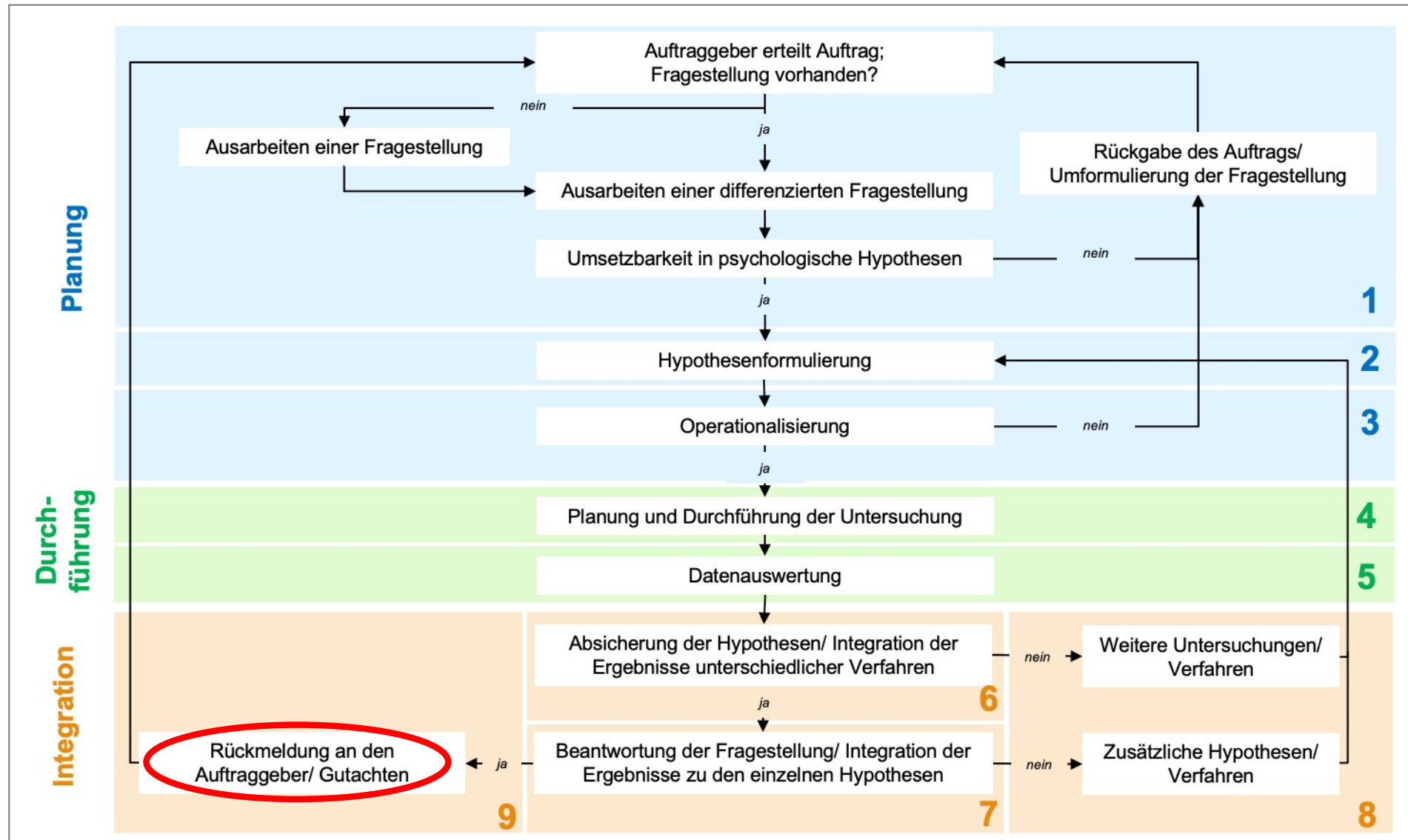


Verteilung von T-Werten und Prozenträngen in einer Normalverteilung

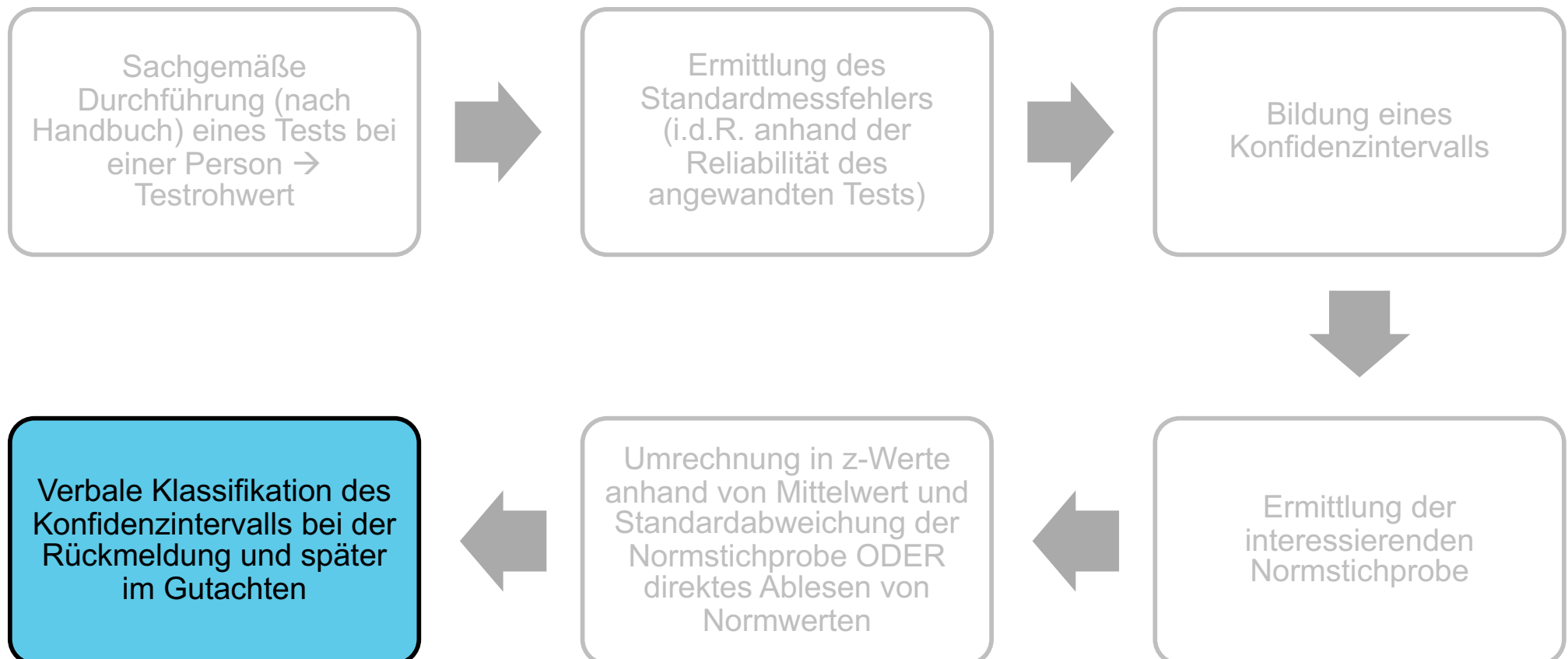
<b>Abstand in Rohwerten</b>	konstant	konstant
<b>Abstand in T-Werten</b>	10	10
<b>Abstand in Prozenträngen</b>	$84 - 50 = 34$	$97,7 - 84 = 13,7$

# 6. Verbalisierung von Konfidenzintervallen & Rückmeldung

## 6. Verbalisierung des Konfidenzintervalls & Rückmeldung

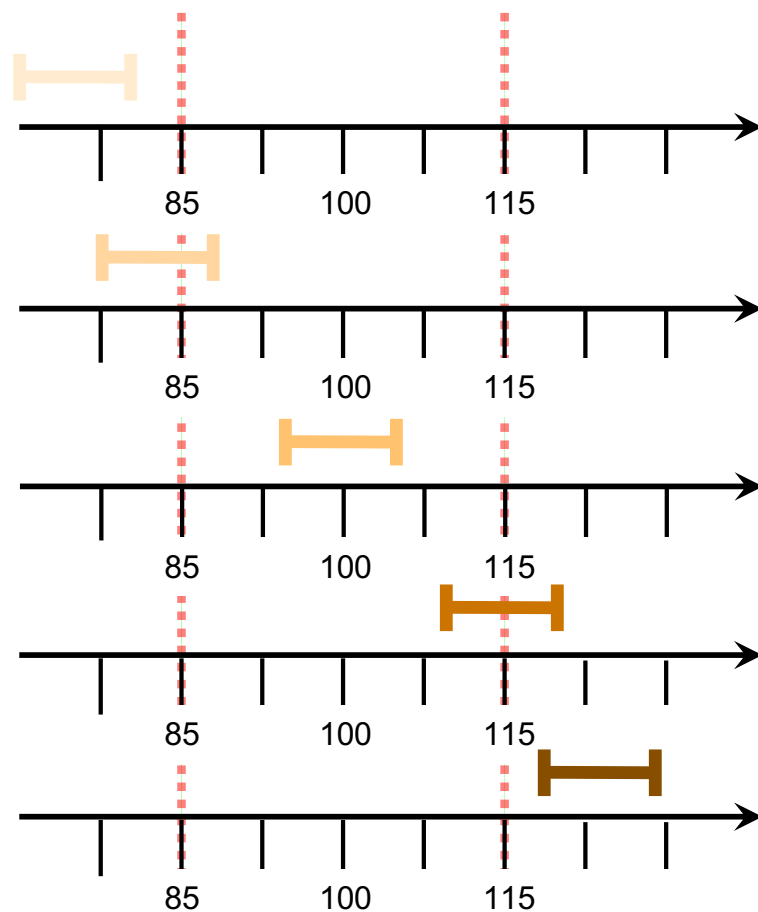


## Schritte der **normorientierten** Einzelfalldiagnostik (am Bsp. eines Tests)



## Klassifikation von Testwerten

(am Beispiel von IQ-Normwerten)



Grenzen des KIs	Beispiel IQ-Werte	Klassifikation des Ergebnisses
< -1SD	< 85	Unterdurchschnittlich
< -1SD bis ≥ -1SD	< 85 bis ≥ 85	Durchschnittlich bis unterdurchschnittlich
≥ -1SD bis ≤ +1SD	≥ 85 bis ≤ 115	Durchschnittlich
≤ +1SD bis > +1SD	≤ 115 bis > 115	Durchschnittlich bis überdurchschnittlich
> +1SD	> 115	Überdurchschnittlich

## **Einflussgrößen auf die Klassifizierung in verbale Kategorien**

### **Wahl des Durchschnittsbereichs:**

- Typischerweise  $MW \pm 1 SD$  (z.B. IQ von 85 bis 115). Bei einer Normalverteilung sind das 68% der Bevölkerung.
- Alternativ: Interquartil-Abstand (50% der Bevölkerung)
- Oder testspezifische Definition: z.B. T = 45 bis 55 im NEO-PI-R (38% der Bevölkerung)

**Wahl des Konfidenzniveaus:** Typischerweise 95% / 90% / 80%

## Mündliche Rückmeldung an die untersuchte Person

Richtet sich i.d.R. an **nicht-fachliche** Personen, und beinhaltet:

- Angabe des Messinstruments mit dem gemessen wurde (z.B. IST-5)
- Inhaltliche Beschreibung des Maßes (z.B. was bedeutet fluide Intelligenz?)
- Hinweis auf Einschränkungen bei der Messgenauigkeit (z.B. „unter Berücksichtigung der Messgenauigkeit“)
- Norm-Wert nicht als Zahl, sondern als verbalisiertes, klassifizierendes Intervall *unter Einbeziehung der Normstichprobe* (z.B. „unterdurchschnittlich bis durchschnittlich im Vergleich zu Gleichaltrigen“)

Achtung: Aufgrund der eingeschränkten Messgenauigkeit lieber Formulierungen wie „Sie schilderten sich als...“ bzw. „Ihre Testleistung war...“, statt „Sie sind...“!

## Schriftliches Gutachten: Ergebnisbericht + Anhang

### Ergebnisbericht: Grundlage für Befund

- Selber Inhalt wie mündliche Rückmeldung, ggf. Teilaspekte etwas ausführlicher
- Zusätzlich: vollständige Zitation des Tests (Autoren, Auflage des Testmanuals)
- Hinweise, unter welchen (ggf. suboptimalen) Bedingungen die Messung stattgefunden hat (je nach Detailgrad auch im Anhang)

## Schriftliches Gutachten: Ergebnisbericht + Anhang

### Anhang (richtet sich an fachliche Leserinnen):

- numerischer Roh- und Normwert und numerische Intervallgrenzen
- verwendetes Konfidenzniveau
- verwendete Berechnungsart (frequentistisch oder bayesianisch)
- beim bayesianischen KI: verwendete Prior (s. Sitzung 9)
- verwendete Reliabilität (oder Standardmessfehler beim modellbasierten KI)
- verwendetes Schema zur verbalen Klassifizierung

# 7. Beispiel

## Use Case 1: KI auf Basis von Rohwerten



Karl K.

- Absolvierte den IST-2000-R
- Testrohwert für schlussfolgerndes Denken: **115**

IST-2000-R

Besitzt Karl im Vergleich zur deutschen  
Bevölkerung überdurchschnittliche Fähigkeiten  
im Bereich schlussfolgerndes Denken?  
(dabei wollen wir uns recht sicher sein können)

Deutsche Gesamtstichprobe:

- N = 3458
- zu viele Gymnasiasten
- keine Angabe zum Alter

Schlussfolgerndes Denken in  
der Gesamtstichprobe:

- Mittelwert: 93.4
- Standardabweichung: **27**

Cronbachs Alpha für die Skala für  
schlussfolgerndes Denken = **.96**

### Schritt 1: KI für Rohwerte berechnen

$$95\% \text{ KI} = x \pm 1.96 * \sigma_x * \sqrt{1 - Rel(X)}$$

$$= 115 \pm 1.96 * 27 * \sqrt{1 - 0.96} = 115 \pm 10.58 = [104.42; 125.58]$$

## Use Case 1: KI auf Basis von Rohwerten



Karl K.

- Absolvierte den IST-2000-R
- Testrohwert für schlussfolgerndes Denken: **115**

IST-2000-R

Besitzt Karl im Vergleich zur deutschen  
Bevölkerung überdurchschnittliche Fähigkeiten  
im Bereich schlussfolgerndes Denken?  
(dabei wollen wir uns recht sicher sein können)

Deutsche Gesamtstichprobe:

- N = 3458
- zu viele Gymnasiasten
- keine Angabe zum Alter

Schlussfolgerndes Denken in  
der Gesamtstichprobe:

- Mittelwert: **93.4**
- Standardabweichung: **27**

Cronbachs Alpha für die Skala für  
schlussfolgerndes Denken = .96

### Schritt 2: KI-Grenzen in Normwerte überführen

95% KI in Rohwerten = **[104.42; 125.58]**

95% KI in z-Werten =  $\left[ \frac{104.42 - 93.4}{27}; \frac{125.58 - 93.4}{27} \right] = \mathbf{[0.41; 1.19]}$

Grundwissen zu Standardwerten

- Mittelwert: 100
- Standardabweichung: 10

95% KI in Standardwerten =  $[100 + (0.41 * 10); 100 + (1.19 * 10)] = \mathbf{[104.10; 111.90]}$

## Use Case 2: KI auf Basis von Normwerten



Karl K.

- Absolvierte den IST-2000-R
- Testwert für schlussfolgerndes Denken: **108 SW (Standardwerte)**

IST-2000-R

Besitzt Karl im Vergleich zur deutschen Bevölkerung überdurchschnittliche Fähigkeiten im Bereich schlussfolgerndes Denken? (dabei wollen wir uns recht sicher sein können)

Grundwissen zu Standardwerten

- Mittelwert: 100
- Standardabweichung: 10

Cronbachs Alpha für die Skala für schlussfolgerndes Denken = .96

$$95\% \text{ KI in Standardwerten} = 108 \pm 1.96 * 10 * \sqrt{1 - 0.96} = 108 \pm 3.92 = [104.10; 111.90]$$

→ Die obere KI-Grenze liegt mehr als eine Standardabweichung über dem Mittelwert der Standardwert-Skala (d.h., > 110). Da die untere KI-Grenze jedoch innerhalb einer Standardabweichung vom Mittelwert liegt, ist das KI „durchschnittlich bis überdurchschnittlich“!

## Use Case: Mündliche Rückmeldung

„Sie haben einen **Test zum schlussfolgernden Denken aus dem Intelligenz-Struktur-Test 2000 R** bearbeitet. Dieser Test misst die **Fähigkeit zum formal-logischen Denken**. **Unter Berücksichtigung der Messgenauigkeit des Tests fällt Ihre Testleistung im Vergleich zu anderen Erwachsenen verschiedener Alters- und Bildungszugehörigkeiten in einen durchschnittlichen bis überdurchschnittlichen Bereich.**“

## Use Case: Ergebnisbericht

Bei einer Intelligenzmessung mit dem **Intelligenz-Struktur-Test 2000 R (Amthauer et al., 2001, 2. Auflage)** ...

...erzielte Herr K. im Bereich des **schlussfolgernden Denkens** einen Wert, ...der **im Vergleich zu anderen Erwachsenen verschiedener Alters- und Bildungszugehörigkeiten** (mit einer Überrepräsentierung von Gymnasiasten im Vergleich zur deutschen Bevölkerung)

...und unter Berücksichtigung der Messgenauigkeit als **durchschnittlich bis überdurchschnittlich** zu klassifizieren ist.

D.h. Herr K. zeigte eine durchschnittlich bis überdurchschnittlich ausgeprägte Fähigkeit zu **formal-logischem Denken anhand sprachlichen, zahlenbezogenen und bildhaft-räumlichen Materials**.

## Use Case: Informationen für den Anhang

- Rohwert mit freq. 95% Konfidenzintervall : 115 [104.42; 125.58]
- Standardwert mit freq. 95% Konfidenzintervall: 108 [104.10; 111.90]
- Reliabilitätsschätzung (Cronbachs Alpha) = .96
- Klassifizierungsschema:

Grenzen des Kls	Klassifikation des Ergebnisses
< -1SD	Unterdurchschnittlich
< -1SD bis $\geq$ -1SD	Durchschnittlich bis unterdurchschnittlich
$\geq$ -1SD bis $\leq$ +1SD	Durchschnittlich
$\leq$ +1SD bis > +1SD	Durchschnittlich bis überdurchschnittlich
> +1SD	Überdurchschnittlich

## Das war's schon fast für heute ...

- *Ausblick:* In der nächsten Vorlesung lernen wir, wie wir in einzelfalldiagnostischen Entscheidungen Vorwissen berücksichtigen können.
- Aber zuerst: Gibt es offene Fragen zur heutigen Vorlesung?