

## Interpretation von Testresultaten und Normierung

Die Interpretation von Testwerten wird mit Hilfe verschiedener Bezugssysteme gemacht, die Maßstäbe und Standards liefern, die helfen einen konkreten Testwert einzuordnen.

### Arten von Bezugssystemen

- a) Normorientiertes Bezugssystem  
Vergleich der Testresultate an der Verteilung der Ergebnisse in einer Bezugsgruppe
- b) Kriteriumsorientiertes Bezugssystem  
Vergleich mit inhaltlich definierten Zielen. Vergleich mit anderen ist irrelevant
- c) Individuelle Bezugsgrößen  
Vergleich vom Erleben und Verhalten einer Person zwischen verschiedenen Zeitpunkten

### a) Normorientierte Testwertinterpretation

= Durch Normierung – auf Basis einer großen und repräsentativen Stichprobe – erhält man ein Bezugssystem, um individuelle Testwerte einordnen bzw. interpretieren zu können.

- Bildung der Norm: Test wird einer möglichst großen und repräsentativen Stichprobe der Zielpopulation, der Normierungs- oder Eichstichprobe vorgelegt.
- Der individuelle Testwert (Rohwert) wird in einen Normwert transformiert, der es ermöglicht den Testwert in Bezug zu der Referenzgruppe zu interpretieren.

#### Warum?

Rohwerte (z.B. Summenscores) sind wenig aussagekräftig!

- Test 1: 80 von 100 Punkten; Test 2: 60 von 100 Punkten
- Der zweite Test kann für eine höhere Leistung/Merkmalsausprägung (relativ zur Referenzgruppe) stehen als der erste Test

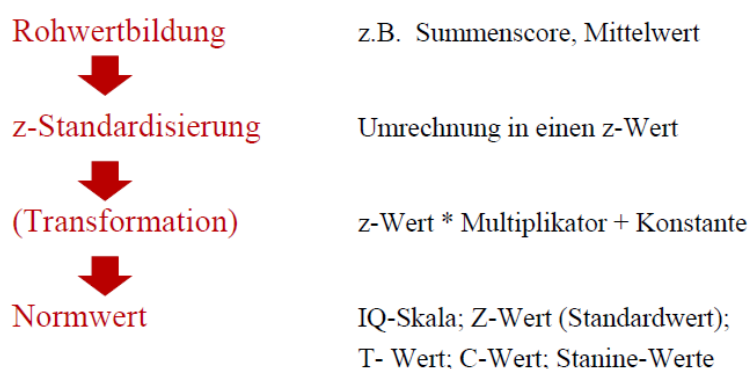
Für die Erstellung von Normen bestehen verschiedene Möglichkeiten:

1. Standardnormen – Abweichungs- und Variabilitätsnormen)
2. Prozentrangnormen
3. Äquivalenznormen (Standardnormäquivalente sowie Alters- und Klassenäquivalente

### 1. Standardnormen

= Abweichungsnormen und Variabilitätsnormen

- Abweichung vom Mittelwert der Vergleichsstichprobe in Einheiten der Standardabweichung
- Voraussetzung: Normalverteilung



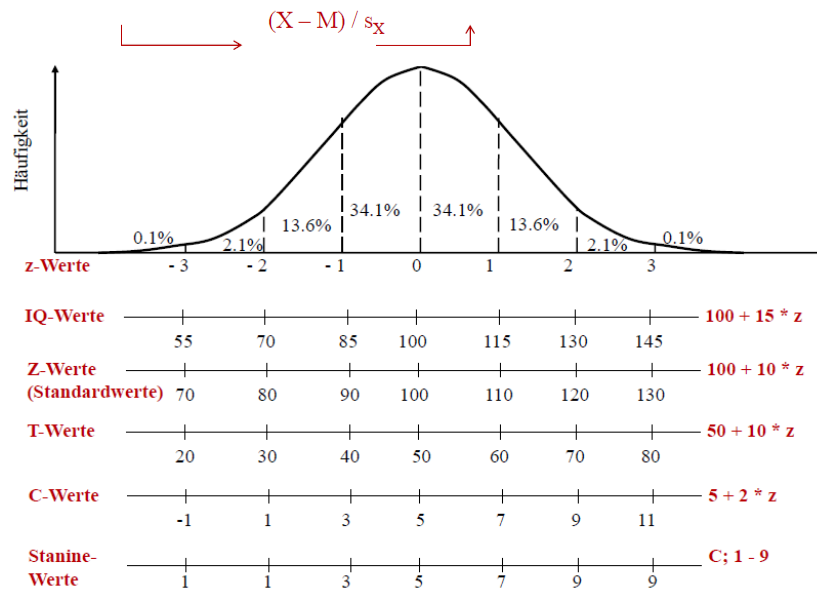
Rohwerte (X):

1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9, 10

(M = 5.35; SD = 2.31) → (M = 0.00; SD = 1.00)

Rohwerteverteilung

Standardnormalverteilung



IQ-Werte

Z-Werte

Intelligenztests

T-Werte

häufig bei Persönlichkeitstests

C-Werte

Stanine-Werte

größere Norm, wenn keine Unterteilung  
unangemessen (z.B. wegen mangelnder  
Reliabilität)

➔ Rohwerte werden durch Normtabellen in die entsprechenden Normwerte umgewandelt (abgelesen)

Was tun, wenn keine Normtabellen im Manual sind?

➔ Wenn Mittelwert und Standardabweichung im Manual angegeben sind, dann kann der Rohwert über die z-Standardisierung in jeden beliebigen Normwert transformiert werden

Wenn ein Merkmal (empirisch) nicht normalverteilt ist, verbietet sich die Transformation in Normwerte (es sei denn, die Annahme für Normalverteilung ist begründbar).

- Rohwerte über Flächentransformation in Normalverteilung überführen
- anschließend Transformation in Normwerte

#### Deskriptive Statistiken

	N	Minimum	Maximum	Mittelwert	Standardabweichung	Schiefe		Kurtosis	
	Statistik	Statistik	Statistik	Statistik	Statistik	Statistik	Standardfehler	Statistik	Standardfehler
Extraversion_Roh	77	11,00	37,00	22,8052	6,01323	-,159	,274	-,576	,541
Neurotizismus_Roh	77	4,00	36,00	16,1299	6,06402	,970	,274	1,150	,541
Gültige Anzahl (listenweise)	77								

Extraversion: Schiefe =  $-.159 / .274 = -.580$

Kurtosis =  $-.576 / .541 = -1.06$  -> Normalverteilung gegeben

Neurotizismus: Schiefe =  $.970 / .274 = 3.54$

Kurtosis =  $1.150 / .541 = 2.12$  -> Normalverteilung nicht gegeben

## 2. Prozentrangnormen

PR oder Perzentile geben an, wieviel Prozent der Referenzgruppe geringere oder maximal gleiche Ausprägungen auf dem gemessenen Merkmal hat

- Rohwerten wird die **relative Position** in einer Bezugsgruppen zugeordnet
- Anteil der Normstichprobe, der gleich große oder kleinere Werte erreicht hat (in Prozent)
- auch bei nicht normalverteilten Rohwerten möglich

**Ermittlung** durch kumulierte prozentuale Häufigkeiten

N = 85 (Testwerte von 5 – 14)

Testwert	5	6	7	8	9	10	11	12	13	14
f	1	0	6	10	19	24	14	6	4	1
cumf	1	1	7	17	36	60	74	80	84	85

$$PR = \frac{cumf}{N} \cdot 100$$

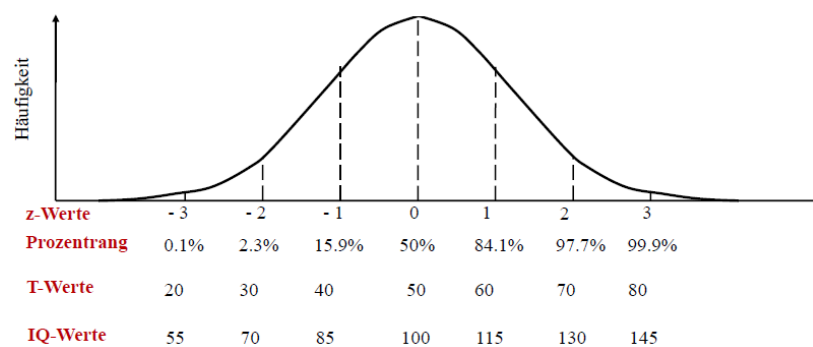
$$PR = 17/85 \cdot 100 = 20 \% \text{ (Prozentrang für Testwert 8)}$$

f = Häufigkeit von Messwerten innerhalb einer Klasse

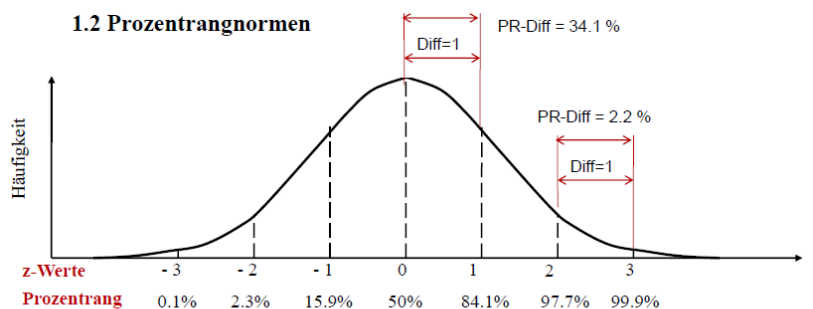
cumf = kumulierte Häufigkeiten der Messwerte bis zur Klassengrenze

N = Gesamtzahl der Probanden

### Zusammenhang zu Normwerten



#### 1.2 Prozentrangnormen



**Problem:** Interpretierbarkeit – Normwerte sind intervallskaliert, Prozenträge hingegen nicht. Daher sind Prozentrangdifferenzen nicht wie Normwertdifferenzen zu interpretieren.

- PR sind einfache verständliche Normen -> gut für Kommunikation mit Personen ohne testtheoretische Vorbildung (z.B. in Gutachten)
- Achtung: Nichtlineare Transformation der Ausgangswerte
  - Skaleneigenschaften der Rohwerte gehen verloren
  - Z-Wert-Differenz von 1 entspricht im Mittelbereich der Skala größeren PR-Differenzen als dies im hohen und niedrigen Bereich der Fall ist (siehe Grafik).
  - Daher keine weitere Berechnung, wie Gruppenmittelwerte, zu empfehlen (-> Rohwerte nehmen)

### 3. Äquivalenznormen

- Zuordnung der Rohwerte zu bestimmten Referenzgruppen z.B. Alter oder Reifeabschnitt, für die die betreffende Leistung besonders typisch ist.
- älteste Art der Normierung
- heutiger Einsatz: Bestimmung des Altersniveau bei z.B. Personen mit geistiger Behinderung

Alter	3	4	5	6	7	8	9	10
Anzahl Aufgaben	12	12	12	12	12	12	12	12
Anzahl gelöst	12	12	12	12	8	5	3	0

Proband: Alter = 7 Jahre 3 Monate (87 Monate)

Intelligenzalter = 6 Jahre + 8 Monate + 5 Monate + 3 Monate  
= 7 Jahre 4 Monate (88 Monate)

$$\text{IQ} = \text{Intelligenzalter} / \text{Alter} = 88/87 = 1.01 * 100 = 101$$

### b) Kriteriumsorientierte Testwertinterpretation

= Interpretation in Bezug auf ein **spezifisches inhaltliches Kriterium**  
(z.B. diagnostische Aussage – Depression ja/nein)

### Bezug des Testwertes auf ein externes Kriterium

## ROC-Analyse zur Bestimmung des Schwellenwertes

➔ Schwellenwert, der ein optimales Gleichgewicht zwischen Sensitivität und Spezifität herstellt.

		Klassifikation	
		+	-
Kriterium	+	RP	FN
	-	FP	RN

RP = richtig positiv

FP = falsch positiv

FN = falsch negativ

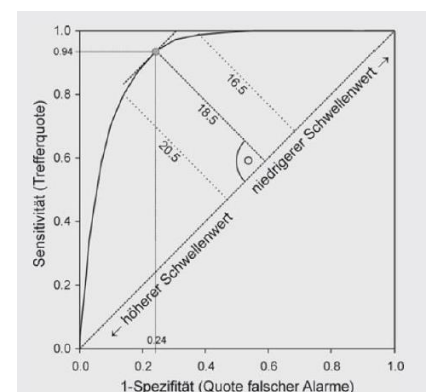
RN = richtig negativ

**Sensitivität** (Trefferquote) =  $RP / FN + RP$

**Spezifität** (Quote korrekter Ablehnungen) =  $RN / FP + RN$

### Beispiel: Depressionsfragebogen

1. Klassifikation: klinisches Interview  
(Depressive vs. nicht Depressive)
2. Erhebung der Testwerten
3. Berechnung der Sensitivität und Spezifität für jeden potentiellen Schwellenwert
4. Optimaler Schwellenwert bei Sensitivität + Spezifität  $\rightarrow$  max



### Bezug des Testwertes auf Aufgabeninhalte

Testwert ist im Sinne des Ausmaßes der Erfüllung eines z.B. Lernziels zu interpretieren.

- Anteil gelöster Aufgaben

### Anwendung bei Lernziel- oder Leistungstests

## c) Testeichung

1. Definition der Zielpopulation
2. Wahl des Erhebungsdesigns

**Globale Repräsentativität:** wenn die Zusammensetzung der Stichprobe hinsichtlich aller möglichen Faktoren mit der Populationszusammensetzung übereinstimmt (Zufallsstichprobe aus der definierten Zielpopulation)

**Spezifische Repräsentativität:** wenn die Zusammensetzung der Stichprobe hinsichtlich derjenigen Faktoren der Populationszusammensetzung repräsentativ ist, die mit dem Testmerkmal in irgendeiner Weise zusammenhängen (z.B. Geschlecht, Alter, Bildungsgrad)