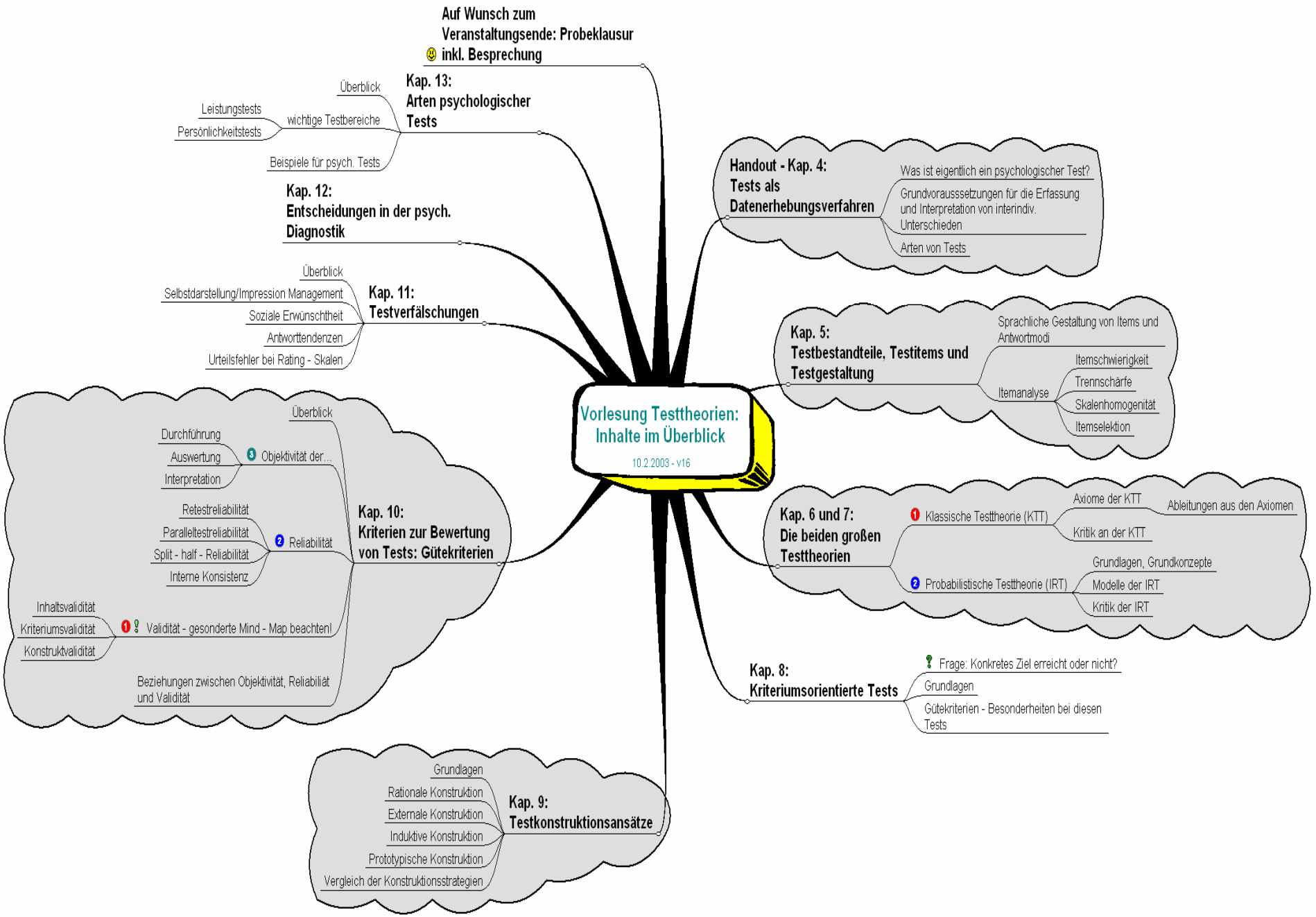
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Vorlesung Testtheorien

Dr. Tobias Constantin Haupt, MBA

Sommersemester 2007


Vorlesung Testtheorien: Inhalte im Überblick
10.2.2003 - v16
Auf Wunsch zum Veranstaltungsende: Probeklausur
inkl. Besprechung
Kap. 13: Arten psychologischer Tests
Überblick
wichtige Testbereiche
Leistungstests
Persönlichkeitstests
Beispiele für psych. Tests
Kap. 12: Entscheidungen in der psych. Diagnostik
Kap. 11: Testverfälschungen
Überblick
Selbstdarstellung/Impression Management
Soziale Erwünschtheit
Antworttendenzen
Urteilsfehler bei Rating - Skalen
Handout - Kap. 4: Tests als Datenerhebungsverfahren
Was ist eigentlich ein psychologischer Test?
Grundvoraussetzungen für die Erfassung und Interpretation von interindiv. Unterschieden
Arten von Tests
Kap. 5: Testbestandteile, Testitems und Testgestaltung
Sprachliche Gestaltung von Items und Antwortmodi
Itemanalyse
Itemschwierigkeit
Trennschärfe
Skalenhomogenität
Itemselektion
Kap. 10: Kriterien zur Bewertung von Tests: Gütekriterien
Überblick
Objektivität der...
Durchführung
Auswertung
Interpretation
Reliabilität
Retestreliabilität
Paralleltestreliabilität
Split - half - Reliabilität
Interne Konsistenz
Validität - gesonderte Mind - Map beachten!
Inhaltsvalidität
Kriteriumsvalidität
Konstruktvalidität
Beziehungen zwischen Objektivität, Reliabilität und Validität
Kap. 6 und 7: Die beiden großen Testtheorien
Klassische Testtheorie (KTT)
Axiome der KTT
Ableitungen aus den Axiomen
Kritik an der KTT
Probabilistische Testtheorie (IRT)
Grundlagen, Grundkonzepte
Modelle der IRT
Kritik der IRT
Kap. 8: Kriteriumsorientierte Tests
Frage: Konkretes Ziel erreicht oder nicht?
Grundlagen
Gütekriterien - Besonderheiten bei diesen Tests
Kap. 9: Testkonstruktionsansätze
Grundlagen
Rationale Konstruktion
Externale Konstruktion
Induktive Konstruktion
Prototypische Konstruktion
Vergleich der Konstruktionsstrategien

4. Axiom

Die Höhe des Messfehlers E ist unabhängig vom Ausprägungsgrad der wahren Werte T' anderer Tests: $\mathbf{r_{T'E} = 0}$.

Beispiel: Die Messfehler eines Intelligenztests sollten z.B. nicht mit Testangst oder Konzentrationsfähigkeit (mit anderen Tests gemessene Persönlichkeitsmerkmale usw.) korrelieren.

5. Axiom

Die Messfehler verschiedener Testanwendungen (z.B. E1 und E2) sind voneinander unabhängig, d.h., ihre Messwerte sind unkorreliert: $r_{E1E2} = 0$.

Beispiel: Personen, die bei einer Testanwendung besonders müde sind oder hohe Testangst haben, sollten bei einer Testwiederholung keine analogen Effekte zeigen.

Ableitungen aus den Axiomen

Reliabilität: Die Reliabilität R gibt den Anteil der Varianz der wahren Werte s_T^2 an der Varianz der beobachteten Werte s_X^2 an:

- **$R = s_{Tt}^2 / s_{Xt}^2$.**

- Diese Aussage eignet sich als Merksatz!

Ableitungen aus den Axiomen

$$R = r_{t,t`} = \frac{s_{wt}^2}{s_{xt}^2} =$$

Die Reliabilität gibt den Anteil der Varianz der wahren Werte an der Varianz der beobachteten Werte an.

Dies ist die wichtigste und zentrale Ableitung aus den Axiomen der klassischen Testtheorie.

Ableitungen aus den Axiomen

$$R = r_{t,t`} = \frac{s_{wt}^2}{s_{xt}^2} = \frac{\text{wahre Varianz}}{\text{Gesamtvarianz}}$$

Die Reliabilität gibt den Anteil der Varianz der wahren Werte an der Varianz der beobachteten Werte an.

Dies ist die wichtigste und zentrale Ableitung aus den Axiomen der klassischen Testtheorie.

Ableitungen aus den Axiomen

Ein Reliabilitätskoeffizient von z. B.
R = .80 gibt an, daß die beobachtete Varianz der Testwerte zu 80 % auf wahre Unterschiede zwischen den Testpersonen zurückzuführen ist und zu 20 % auf Fehlervarianz beruht.

- Dazu rechnen wir am besten gleich eine kleine Übungsaufgabe.

Ableitungen aus den Axiomen

Pbn	Messwert x	Wahrer Wert Tx	Messfehler
1	2	1	1
2	0	1	-1
3	5	5	0
4	10	9	1
5	8	9	-1

- Prüfen Sie, ob für den Messfehler der Messwerte x die beiden Axiome der KTT (= Messfehlertheorie) gelten.
- Berechnen Sie die Reliabilität

	2. Axiom $\mu(E) = 0$ 3. Axiom $r_{er}=0$

Ableitungen aus den Axiomen

Pbn	Messwert x	Wahrer Wert Tx	Messfehler
1	2	1	1
2	0	1	-1
3	5	5	0
4	10	9	1
5	8	9	-1

- Prüfen Sie, ob für den Messfehler der Messwerte x die beiden Axiome der KTT (= Messfehlertheorie) gelten.
- Berechnen Sie die Reliabilität

$$R = \frac{s_w^2}{s_x^2} =$$

2. Axiom
$\mu(E) = 0$

3. Axiom
$r_{er}=0$

Ableitungen aus den Axiomen

Pbn	Messwert x	Wahrer Wert Tx	Messfehler
1	2	1	1
2	0	1	-1
3	5	5	0
4	10	9	1
5	8	9	-1

- Prüfen Sie, ob für den Messfehler der Messwerte x die beiden Axiome der KTT (= Messfehlertheorie) gelten.
- Berechnen Sie die Reliabilität

$$R = \frac{s_w^2}{s_x^2} = \frac{16}{17} = 0{,}94$$

2. Axiom
$\mu(E) = 0$

3. Axiom
$r_{er}=0$

Ableitungen aus den Axiomen

Die KTT wird oft auch als Messfehlertheorie bezeichnet. Deshalb wollen wir uns diesem Messfehler einmal näher widmen.

Ganz wichtig ist die Grundüberlegung, psychologische Diagnostik und Tests nicht (verantwortlicherweise) ohne Betrachtung des Messfehlers betreiben zu können.

Eine hohe Reliabilität (und damit ein kleiner Standardmessfehler) ist in der Praxis sehr wichtig, da dies die „Breite" der zu bestimmenden Konfidenzintervalle wesentlich mitbestimmt.

Ableitungen aus den Axiomen

Der *Standardmessfehler* (in drei Formulierungen)

- ist derjenige Anteil an der Streuung eines Tests, der zu Lasten seiner (gewöhnlich nicht perfekten, also „unvollständigen“) Reliabilität geht
- ist ein Maß für den Anteil der Fehlerstreuung an der Streuung von Messwerten
- gibt die Streuung der beobachteten Werte um die entsprechenden wahren Werte bei Messwiederholungen einer Person an (läßt sich als Normalverteilung mit wahrem Wert als Zentrum veranschaulichen).

Ableitungen aus den Axiomen

Der Standardmessfehler berechnet sich nach

$$s_e = s_x * \sqrt{(1 - R)}$$

und hängt somit von der Streuung s und dem Reliabilitätskoeffizienten R ab
(bei perfekter Reliabilität beträgt er 0; bei fehlender Reliabilität entspricht er der Streuung der beobachteten Werte, welche dann ausschließlich auf Fehlereinflüssen beruhen)

➞ je reliabler das Messinstrument, desto geringer der Standardmessfehler)

Standardmessfehler und Konfidenzintervall gehören eng zusammen (im wahrsten Sinne..)!

Beispiel einer Fragestellung:

Gegeben einen beobachteten Wert X einer Person, in welchem Bereich liegt der wahre Wert W mit einer bestimmten Wahrscheinlichkeit (meist 95%)?

Je geringer der Standardmessfehler, desto schmaler dieser Bereich. Das Konfidenzintervall berechnet sich folgendermaßen:

beobachteter Wert ± Irrtumswahrscheinlichkeits-z-Wert mal Standardmessfehler

Konfidenzintervalle

Man schreibt:

$$x - z_{1-\alpha/2} * s_e \leq w \leq x + z_{1-\alpha/2} * s_e$$

Die *untere* Grenze des Konfidenzintervalls wird also berechnet durch

$$x - z_{1-\alpha/2} * s_e$$

die *obere* Grenze entsprechend durch

$$x + z_{1-\alpha/2} * s_e$$

Konfidenzintervalle

Wenn jemand 110 Punkte im Test erzielt hat und der Standardmessfehler $s_e = 2$ ist, dann liegt der wahre Wert der Person bzgl. ihres IQ mit einer Wahrscheinlichkeit von 95 % zwischen 106,08 und 113,92.

Man schreibt dann:

$$110 - (1{,}96 * 2) \leq w \leq 110 + (1{,}96 * 2)$$

$$106{,}08 \leq w \leq 113{,}92$$

Das heißt, daß z. B. die Aussage, daß die Person einen höheren IQ als 105 hat, auf dem 5-%-Niveau signifikant ist, da das komplette Konfidenzintervall oberhalb von 105 liegt.

Sie erinnern sich an unser Beispiel?

$106{,}08 \leq w \leq 113{,}92$

Konfidenzintervall für den wahren Wert

Das Konfidenzintervall für den wahren Wert einer Person in einem Test wird also in zwei Schritten berechnet:

1. <u>Schritt:</u> Berechnung des Standardmessfehlers s_e

- $S_e = S_x \sqrt{1 - R_X}$

S_x = Standardabweichung des Testwertes X
R_X = Reliabilität des Tests

Konfidenzintervall für den wahren Wert

2. Schritt: Schätzung für das Konfidenzintervall für den unbekannten, wahren Wert w.
Das 95%-Konfidenzintervall ergibt sich unter der Annahme, dass E normalverteilt ist, aus:

- $x - (1{,}96 * Se) \leq w \leq x + (1{,}96 * Se)$

Beispielrechnung die zweite…

$X = 112$

$Sx = 12$

$R_X = 0.84$

$$Se = Sx \sqrt{1 - R_X}$$

$$= 12 \sqrt{1 - 0.84}$$

$$= 12 \bullet 0.4 = 4.8$$

$$x - (1{,}96 * Se) \leq w \leq x + (1{,}96 * Se)$$

$$112 - (1.96 \bullet 4.8) \leq w \leq 112 + (1.96 \bullet 4.8)$$

$$102.6 \leq w \leq 121.4$$

Zwischen diesen beiden Grenzen liegt bei 5% Irrtumswahrscheinlichkeit der wahre Wert der Person.

Reliabilität und Testlänge

- Es läßt sich zeigen, daß z.B. mit der Verdopplung der Testlänge/der Itemanzahl (in Einheiten von homogenen bzw. äquivalenten Aufgaben!) eine Vervierfachung der wahren Varianz einhergeht, während sich die Fehlervarianz nur verdoppelt.

- Da Reliabilität als Anteil der wahren Varianz an der Gesamtvarianz definiert ist, würde dies eine Verdoppelung der Reliabilität bedeuten. Diese mathematische Ableitung hat sich auch empirisch gut bestätigen lassen, was für eine Angemessenheit der Axiome der KTT spricht.

Spearman-Brown-Formel

Der Zusammenhang zwischen Ausgangsreliabilität, Testverlängerung (Faktor k, manchmal auch n abgekürzt; bedeutungsmäßig ist das egal) und neuer Reliabilität läßt sich wie folgt berechnen:

$$r_{tt}neu = k \cdot \frac{r_{tt}}{1+(k-1)\cdot r_{tt}}$$

Dabei zeigt sich, daß der Reliabilitätszuwachs um so größer ist, je geringer die Ausgangsreliabilität ist.

Reliabilität und Testlänge

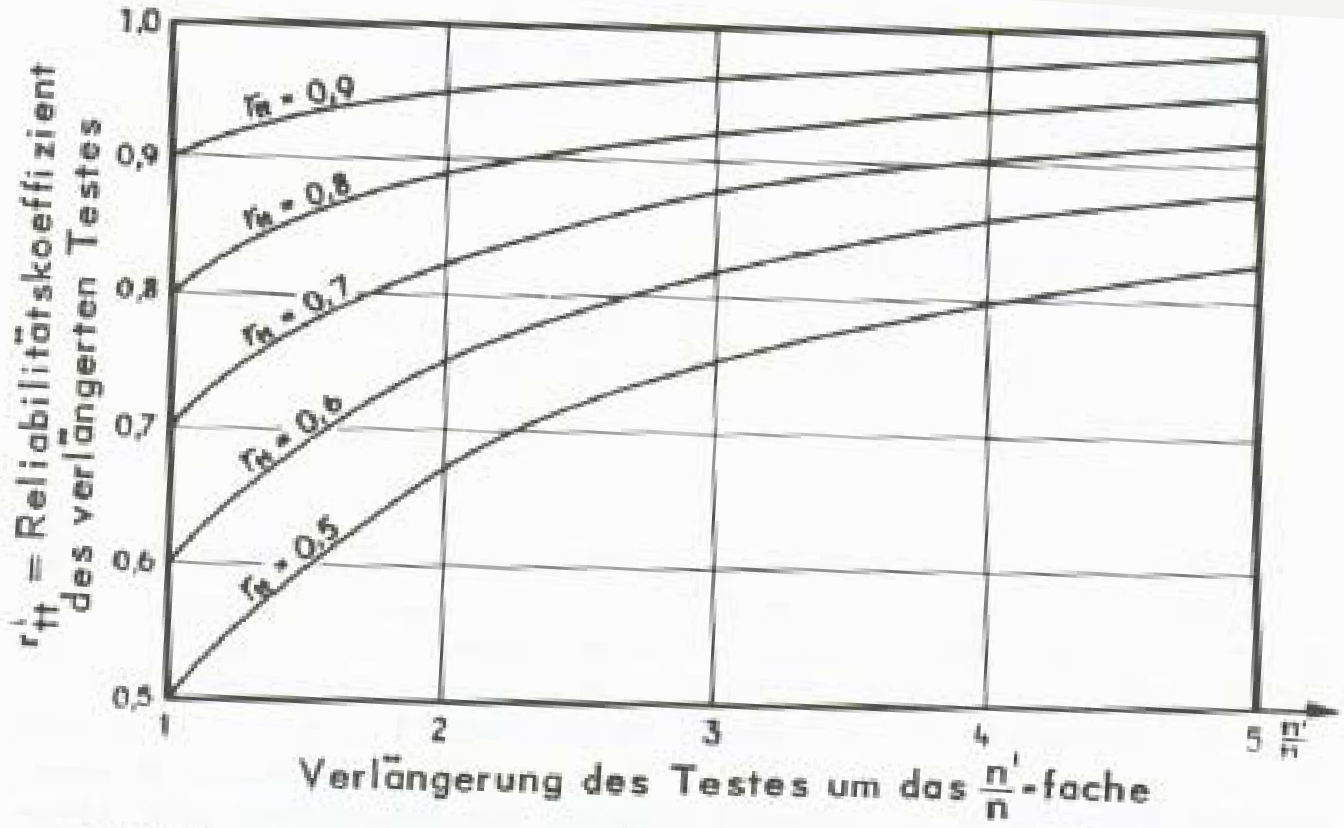


Abb. 10.3 Nomogramm zur Aufwertung eines Reliabilitätskoeffizienten bei Vervielfachung der Testlänge gemäß der verallgemeinerten SPEARMAN-BROWNschen Formel (10.50).

„Kritische“ Differenzen

Problemstellung:

- *Sind beobachtete Skalenwertdifferenzen statistisch signifikant?*
- Unterscheiden sich die Testwerte eines Pbn auf zwei Skalen signifikant (**intra**individuell)?
- Unterscheiden sich die Skalenwerte von zwei Pbn auf derselben Skala signifikant (**inter**individueller Vergleich)?

Kritische Differenz

- Eine Skalenwertdifferenz ist dann **signifikant**, wenn sie **größer** oder gleich **der kritischen Differenz** ist.
- Das wollen wir anhand von Beispielen berechen…

Ermittlung von Unterschieden zwischen zwei Testpunktwerten

Um zu ermitteln, ob sich die Testwerte zweier Probanden in einem Test überzufällig voneinander unterscheiden oder durch Zufallseinflüsse (aufgrund von Unreliabilität des Tests) zu erklären sind, läßt sich eine kritische Differenz berechnen, die empirisch zu übertreffen ist, um von einem sign. Unterschied auszugehen:

$$D_{Krit} = z_{\alpha/2} \cdot s_{eDiff} \qquad s_{eDiff} = s_x \sqrt{2(1 - r_{tt})}$$

Beispiel: Kritische Differenz

Ein Test wurde auf eine Standardabweichung von **20** normiert. Die Reliabilität dieses Tests beträgt **0.92**.
Der **Standardfehler** einer **inter**individuellen Differenz für diesen Test beträgt:

s_e
(bitte berechen Sie diese Größe!)

$$s_{eDiff} = s_x \sqrt{2(1-r_{tt})}$$

Kritische Differenzen

$D_{Krit} = 8 * 1.96 = \underline{15.7}$ *(für $\alpha = .05$)*

Unterschiede zw. Testergebnissen

- Weiteres Übungsbeispiel:
  IQ-Test mit Standardabweichung s_x=10, Reliabilität r_{tt}=.80 und 5%-Irrtumswahrscheinlichkeit (z=1,96).
  D_{krit} beträgt dann… ?
  Zwei Probanden müssen sich in diesem Test also um ? Punkte unterscheiden, um von einer sign. Differenz (die mit 5%-iger Irrtumswahrscheinlichkeit nicht durch die Unreliabilität des Tests zu erklären ist) zwischen beiden Punktwerten ausgehen zu können.

$$D_{Krit} = z_{\alpha/2} \cdot s_{eDiff} \qquad s_{eDiff} = s_x \sqrt{2(1-r_{tt})}$$

Kritik an der KTT

Über 95 Prozent der auf dem Markt befindlichen Testverfahren wurden nach der KTT konstruiert .

.. sie haben sich in der Praxis eindeutig bewährt

Trotzdem sollten auch die „Unzulänglichkeiten“ kritisch betrachtet werden.

Daher ist Folgendes notwendig:

Kritik an der Klassischen Testtheorie

Kritik an der KTT

- Stichprobenabhängigkeit der Parameter und Gütekriterien
- Messtheoretische Probleme
- Wissenschaftstheoretisch fundierte Probleme
- ... aber eins nach dem anderen:

Stichprobenabhängigkeit

- Item- und Testkennwerte (Schwierigkeit, Trennschärfe, Reliabilität, Validität ...) werden an spezifischen Stichproben berechnet
  → Sind diese Befunde **generalisierbar**?
- Man kann z.B. durch die Wahl heterogener oder homogener Stichproben die Reliabilität künstlich erhöhen oder senken.

Stichprobenabhängigkeit

- *Homogenität und Heterogenität*:
  Je homogener eine Stichprobe ist, desto geringer fallen die jeweiligen Korrelationen aus. Dies führt zu einer Varianz der Reliabilitätskoeffizienten, die allein auf die Auswahl der Stichprobe zurückzuführen ist. Reliabilitäten sind somit nur schwer zu generalisieren.

Kritik an der KTT

- Bedeutung der **Stichprobenrepräsentativität**
- Problem der **Definition** der Population

Messtheorie

- Daten sollten auf **Intervallskalenniveau** liegen…Bei vielen Tests ist jedoch fraglich, ob diese Voraussetzung erfüllt ist (so müssten etwa die Abstände bei abgestuften Rating-Skalen psychologisch gleich interpretiert werden: Äquidistanz),

- Berechnung von ***Mittelwerten*** und ***Varianzen***
- Bildung von ***Messwertdifferenzen***
- Das ist **fraglich** und es gibt **keine** explizite **Überprüfung**.

Wissenschaftstheorie

- ***Axiomatische Fehlertheorie*** ohne psychologische Fundierung mit ***nicht überprüfbarer Axiomatik***.

Problematische Schlussfolgerungen aus den Axiomen

- ***Intuitiv problematische Axiome***: z.B.

Axiom 2: Die Annahme, daß der Messfehler bei Messwiederholungen einen Erwartungswert von 0 hat ist bestenfalls eine theoretische Modellannahme.

Axiom 5: Es ist unplausibel anzunehmen, daß die Ausprägung der Messfehler von zwei Tests unabhängig voneinander sind, da sich z.B. Testangst bei beiden systematisch in die gleiche Richtung auswirken könnte.

Kritik an der KTT

Wahrscheinlichkeitsaussagen lassen sich (streng genommen) nicht auf den Einzelfall übertragen:

- Bei Reliabilitäten und Validitäten < 1.0 können lediglich gruppenstatistische Wahrscheinlichkeitsaussagen gemacht werden

- Dies steht der Anwendung psychologischer Tests auf den Einzelfall diametral entgegen.

Kritik an der KTT

Praktische Bewährung:

Nach der KTT entwickelte Verfahren haben sich in der Praxis zur Bestimmung von intra- und interindividuellen Unterschieden bewährt und erlauben objektive, reliable und relativ hoch valide Einschätzungen. Nach wie vor werden fast alle der erhältlichen Testverfahren nach dem Itemanalyseverfahren der KTT entwickelt.

Gute Alternative zu Zufallsentscheidungen:

Instrumentarien der KTT sind immerhin wesentlich besser als Zufallsentscheidungen.

Kritik an der KTT

Auf Grundlage der Kritik an der KTT entwickelten sich

- **Erweiterungen** des ursprünglichen Ansatzes (z. B. Generalisierbarkeitstheorie)
- **Probabilistische Testtheorien**
- Modelle mit **anderer Akzent**setzung (kriteriumsorientierte Leistungsmessung)

Reminder: Gegenstand der Testtheorien

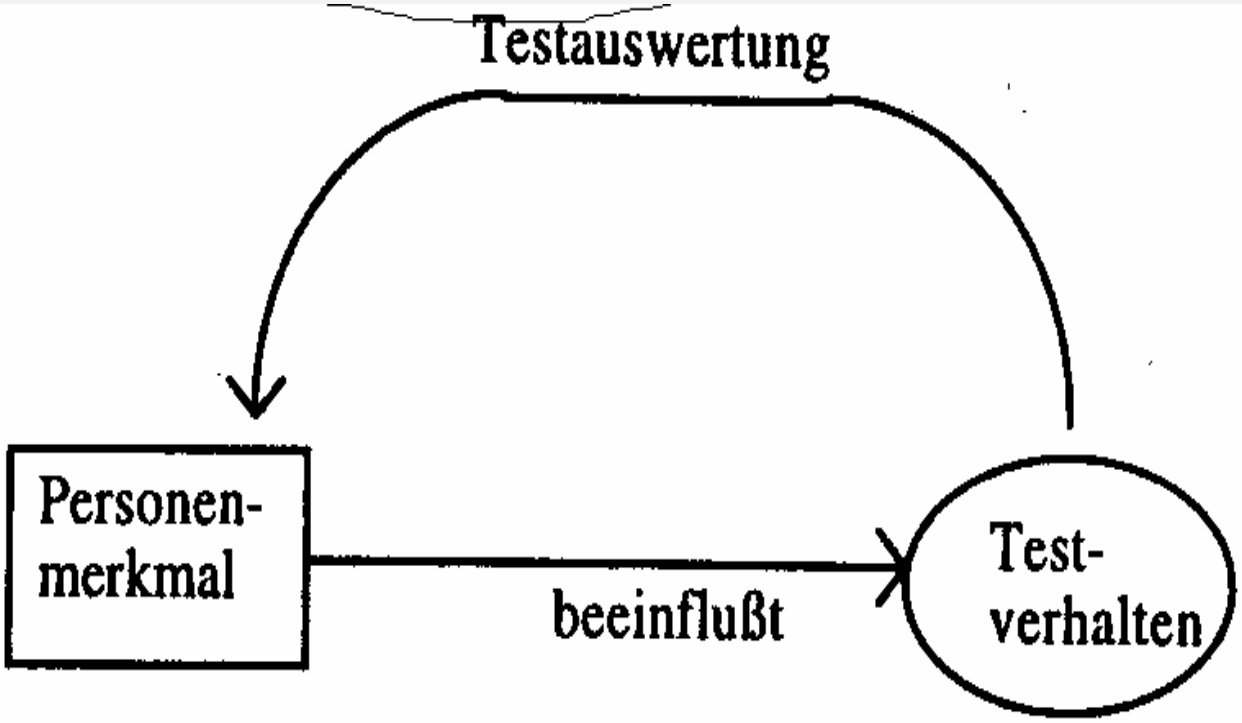


Abbildung 1: Der Gegenstandsbereich der Test-theorie

Auch bekannt unter der Bezeichnung: **Item-Response-Theorie (IRT)**

Ausgangspunkt:

- Ein Test liefert ein Resultat oder ein Testergebnis.
- Dieses Testergebnis soll indikativ für ein best. Merkmal der getesteten Person sein.
- Es ist klar, daß das Testergebnis kein unfehlbares numerisches Äquivalent für die entsprechende Merkmalsausprägung ist (Sie erinnern sich an den Messfehler aus der KTT? „Wer misst, misst… na Sie wissen schon).

Über den Zusammenhang zwischen Testergebnis und Merkmalsausprägungen werden unterschiedliche Annahmen gemacht:

- Die *klassische Testtheorie (KTT)* geht davon aus, daß das Testergebnis direkt (wenn auch mit Messfehlern behaftet) dem Ausprägungsgrad des gemessenen (tatsächlichen, wahren) Merkmals entspricht.
- Der Zusammenhang zwischen Personmerkmal und Testergebnis wird also a priori als **deterministisch** angenommen und ist zudem (weil axiomatisch) keiner empirischen Überprüfung zugänglich.

Testergebnis = Wahrer Wert + Messfehler

Der Effekt unkontrollierter Variablen wird als *Messfehler* bezeichnet

Demgegenüber legt die

- *probabilistische Testtheorie (IRT)* nicht von vornherein fest, wie der Zusammenhang zwischen Merkmalsausprägung und Testergebnis zu sein hat

- Vielmehr unterscheidet sie explizit zwischen Merkmalsebene (latente Variablen) und Testebene (Itemebene; manifeste Variablen) und betrachtet das Testergebnis lediglich als Indikator für das entsprechende Merkmal

- Dabei ist die Beziehung zwischen Merkmal und Indikator (meist als Funktion ausgedrückt) in der Regel eine probabilistische (deterministisch kann sie dabei im Extremfall sein) ist, deren Verlauf zudem sehr unterschiedlich sein kann.

- Das Hauptunterscheidungsmerkmal zur KTT besteht jedoch darin, daß bei der IRT eine (hypothetisch) festgelegte Funktionsform empirisch darauf geprüft werden kann, ob sie auch tatsächlich vorliegt.

Unterschiede zur klassischen Theorie

Klassische Theorie

beginnt mit Annahmen über Tests und führt Items erst bei der Konstruktion reliabler und valider Verfahren ein

Probabilistische Theorie

startet mit Annahmen über Items, aus denen dann Eigenschaften weiterer Testmerkmale (Homogenität) abgeleitet werden

Unterschiede zur klassischen Theorie

Klassische Theorie

beginnt mit Annahmen über Tests und führt Items erst bei der Konstruktion reliabler und valider Verfahren ein

Klassische Theorie

siedelt Testwerte und wahre Werte auf dem gleichen Kontinuum an

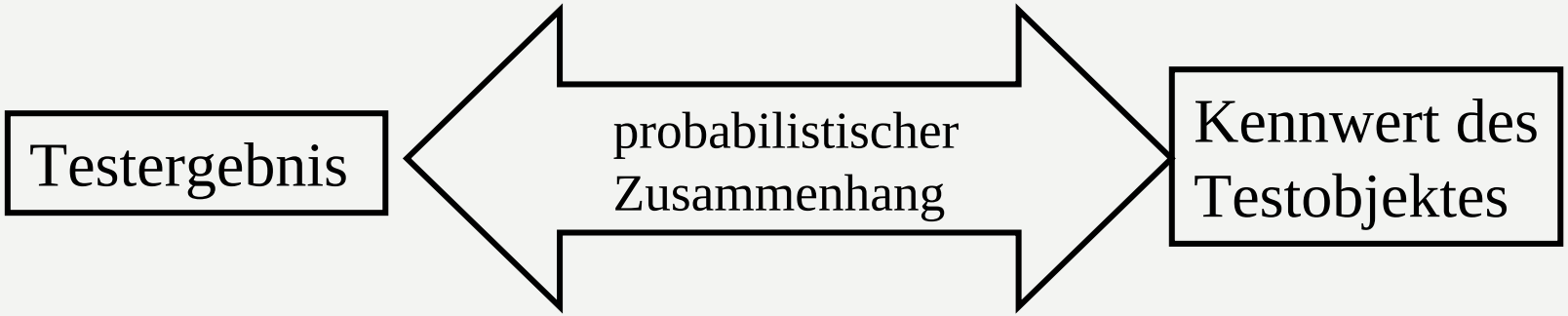
Probabilistische Theorie

startet mit Annahmen über Items, aus denen dann Eigenschaften weiterer Testmerkmale (Homogenität) abgeleitet werden

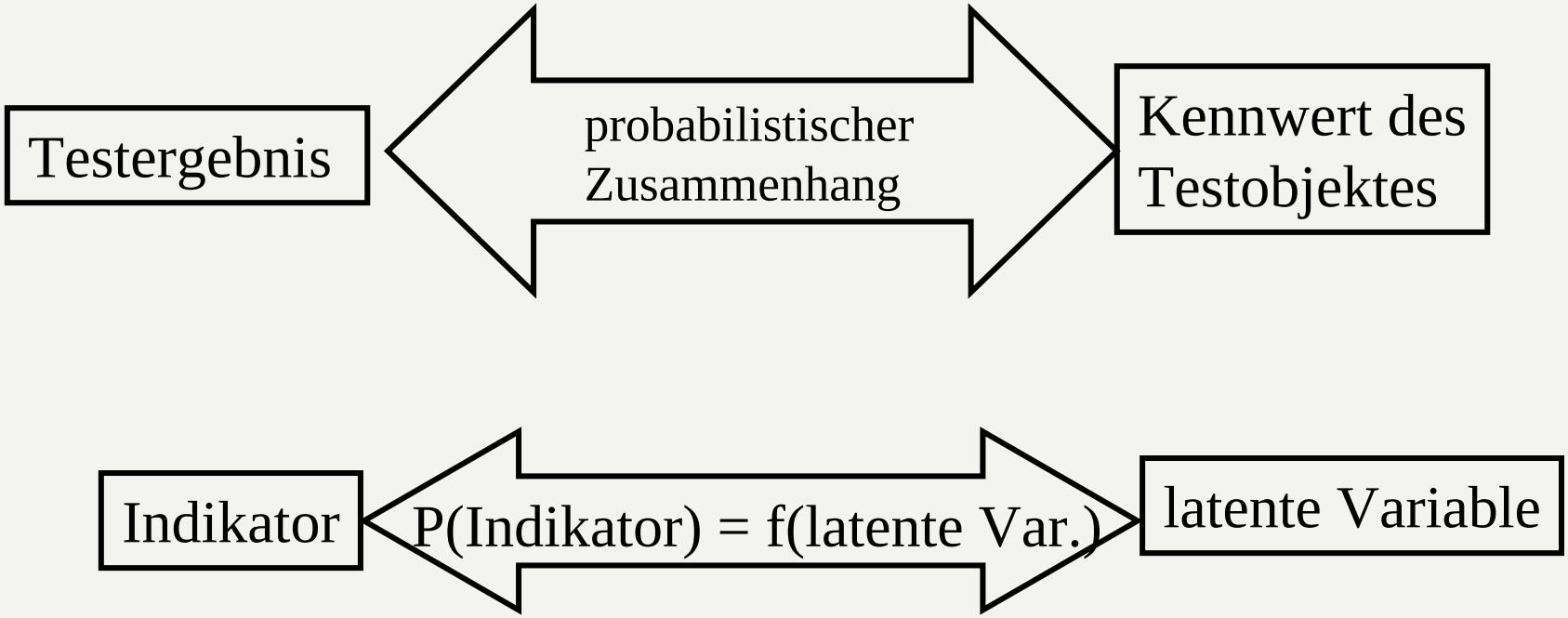
Probabilistische Theorie

betrachtet zwei verschiedene Arten von Variablen, zwischen denen ein probabilistischer Zusammenhang besteht

Fragestellung in der IRT: Welche Rückschlüsse können auf Personenmerkmale gezogen werden, wenn *lediglich* Antworten auf diverse Testitems (item-responses) vorliegen?



Fragestellung in der IRT: Welche Rückschlüsse können auf Personenmerkmale gezogen werden, wenn *lediglich* Antworten auf diverse Testitems (item-responses) vorliegen?



Fragestellung in der IRT: Welche Rückschlüsse können auf Personenmerkmale gezogen werden, wenn *lediglich* Antworten auf diverse Testitems (item-responses) vorliegen?

Variablenebenen in der IRT

manifeste Variablen:

beschreiben das (unterschiedliche) Antwortverhalten auf verschiedene Testitems

latente Variablen ξ (Xi):

bezeichnen die nicht-beobachtbaren Merkmalsausprägungen (Fähigkeiten, Dispositionen), die dem manifesten Verhalten zugrunde liegen sollen

Modellparameter in der IRT

Fähigkeitsparameter (Personenparameter, Dispositionsparameter, ξ oder β (aβility)):

Beschreibt die ***Fähigkeit*** einer Person (Merkmalsausprägung des latenten Traits), ein best. Testitem zu lösen.

Schwierigkeitsparameter (Itemparameter, Anforderungsparameter, σ oder δ (δifficulty):

Anforderung, welche ein Item an die Fähigkeit der zu untersuchenden Person stellt.

Modellparameter in der IRT

Diskriminationsparameter (λ):

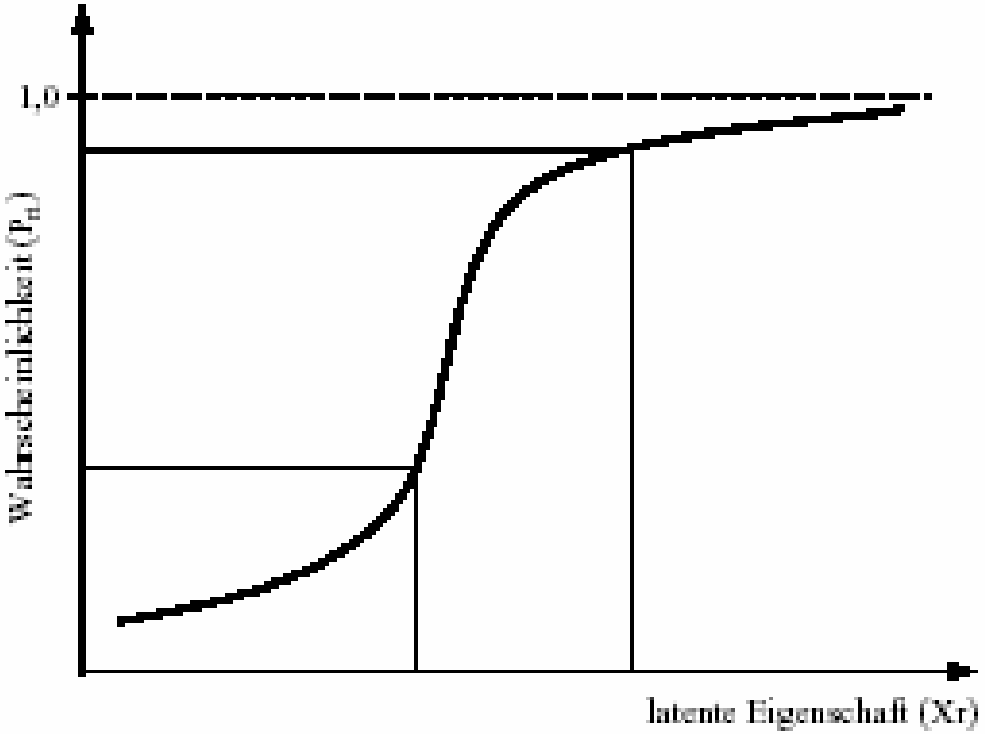
Determiniert die Steilheit der IC-Funktion. Wird nicht in allen Modellen angenommen (z.B. weder im Guttman-Modell, noch in dichotomen Rasch-Modellen).

Zusammenhänge der Modellparameter:

Personen- und Itemparameter lassen sich gemeinsam auf einer eindimensionalen Skala abbilden (joint scale), so daß immer entscheidbar ist, welcher der beiden Parameter größer ist.

Von der Ausprägung beider Parameter soll nun wiederum probabilistisch abhängen, ob ein Item gelöst wird oder nicht, d.h., daß jeder Parameterkonstellation ein best. Wahrscheinlichkeitswert zugeordnet werden kann, mit dem ein Item gelöst wird.

Probabilistische Testtheorie



$P_{ri} = f(D_i; X_r)$ Wahrscheinlichkeit, daß Person r das Item i löst

X_r … Personen-/Fähigkeitsparameter

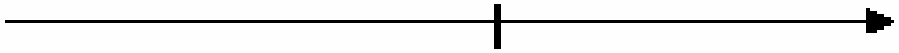
D_i … Itemparameter (z.B. Schwierigkeit oder Leichtigkeit eines Items)

Je höher die Fähigkeit, desto größer ist die Wahrscheinlichkeit, das Item zu lösen!

Abbildung 6-1: *Veranschaulichung* einer Zuordnung von Personen- und Itemparameter (PP, IP) auf einer eindimensionalen Skala.

Drei Fälle:

1. Fall: Die Wahrscheinlichkeit, daß die Person j das Item i löst, ist *gleich 0.50.*



PP = IP

2. Fall: Die Wahrscheinlichkeit, daß die Person j das Item i löst, ist *größer als 0.50.*

IP < PP

3. Fall: Die Wahrscheinlichkeit, daß die Person j das Item i löst, ist *kleiner als 0.50.*

PP < IP

(Fischer, 1990)